cited in the European Search Report of EP-97906876.4 Your Ref.: 090101 012 EP1

WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau



PCT

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6:

C07H 21/04, C12N 1/20, 1/14, 5/00, 9/38, 9/42, C08B 30/04

(11) International Publication Number:

WO 98/24799

A1

(43) International Publication Date:

11 June 1998 (11.06.98)

(21) International Application Number:

PCT/US97/22623

(22) International Filing Date:

8 December 1997 (08.12.97)

(30) Priority Data:

60/056,916 Not furnished 6 December 1996 (06.12.96)

US 10 October 1997 (10.10.97) US

(71) Applicant (for all designated States except US): DIVERSA CORPORATION [US/US]; 10665 Sorrento Valley Road, San Diego, CA 92121 (US).

(72) Inventors; and

- (75) Inventors/Applicants (for US only): BYLINA, Edward, J. [US/US]; Apartment A-1, West Court, Andalusia, PA 19020 (US). SWANSON, Ronald, V. [US/US]; Apartment A, 309 No. Lemon Street, Media, PA 19063 (US). MATHUR, Eric, J. [US/US]; 2654 Galicia Way, Carlsbad, CA 92009 (US). LAM, David, E. [US/US]; 1518 West 249th Street, Harbor City, CA 90710 (US).
- (74) Agent: HAILE, Lisa, A.; Fish & Richardson P.C., Suite 1400, 4225 Executive Square, La Jolla, CA 92037 (US).

(81) Designated States: AU, CA, IL, JP, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: GLYCOSIDASE ENZYMES

(57) Abstract

Thermostable glycosidase enzymes derived from various Thermococcus, Staphylothermus and Pyrococcus organisms is disclosed. The enzymes are produced from native or recombinant host cells and can be utilized in the food processing industry, pharmaceutical industry and in the textile industry, detergent industry and in the baking industry.

cried in the Harachien States of the Roper of the Your Pref.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	
AT	Austria	FR	France	LU	Luxembourg	SN	Slovakia
ΑU	Australia	GA	Gabon	LV	Latvia		Senegal
ΑZ	Azerbaijan	GB	United Kingdom	MC	Monaco	SZ	Swaziland
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TD TG	Chad
BB	Barbados	GH	Ghana	MG	Madagascar		Togo
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TJ	Tajikistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TM	Turkmenistan
BG	Bulgaria	HU	Hungary	ML	Mali	TR	Turkey
BJ	Benin	IE	Ireland	MN	Mongolia	TT	Trinidad and Tobago
BR	Brazil	IL	İsrael	MR	Mongona Mauritania	UA	Ukraine
BY	Belarus	IS	Iceland	MW	Malawi	UG	Uganda
CA	Canada	IT	Italy	MX	Maiawi Mexico	US	United States of Americ
CF	Central African Republic	JP	Japan	· NE	Niger	UZ	Uzbekistan
CG	Congo	KE	Kenya	NL	Netherlands	VN	Viet Nam
CH	Switzerland	KG	Kyrgyzstan	NO		YU	Yugosłavia
CI	Côte d'Ivoire	KP	Democratic People's	NZ	Norway New Zealand	zw	Zimbabwe
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT			
CU	Cuba	KZ	Kazakstan	RO	Portugal		
CZ	Czech Republic	LC	Saint Lucia		Romania		
DE	Germany	Li	Liechtenstein	RU	Russian Federation		
DK	Denmark	LK	Sri Lanka	SD	Sudan		•
EE	Estonia	LR	Liberia	SE	Sweden		
		LI	LIDEI IN	SG	Singapore		

GLYCOSIDASE ENZYMES

BACKGROUND OF THE INVENTION

1. Field of the Inventions

This invention relates to newly identified polynucleotides, polypeptides encoded by such polynucleotides, the use of such polynucleotides and polypeptides, as well as the production and isolation of such polynucleotides and polypeptides. More particularly, the polynucleotides and polypeptides of the present invention has been putatively identified as glucosidases, α -galactosidases, β -galactosidases, β -mannosidases, β -mannases, endoglucanases, and pullalanases.

2. Description of Related Art

The glycosidic bond of β -galactosides can be cleaved by different classes of enzymes: (i) phospho-β-galactosidases (EC3.2.1.85) are specific for a phosphorylated substrate generated via phosphoenolpyruvate phosphotransferase system (PTS)-dependent uptake; (ii) typical β-galactosidases (EC 3.2.1.23), represented by the Escherichia coli LacZ enzyme, which are relatively specific for β -galactosides; and (iii) β -glucosidases (EC 3.2.1.21) such as the enzymes of Agrobacterium faecalis, Clostridium thermocellum, Pyrococcus furiosus or Sulfolobus solfataricus (Day, A.G. and Withers, S.G., (1986) Purification and characterization of a β-glucosidase from Alcaligenes faecalis. Can. J. Biochem. Cell. Biol. 64, 914-922; Kengen, S.W.M., et al. (1993) Eur. J. Biochem., 213, 305-312; Ait, N., Cruezet, N. and Cattaneo, J. (1982) Properties of β-glucosidase purified from Clostridium thermocellum. J. Gen. Microbiol. 128, 569-577; Grogan, D.W. (1991) Evidence that β-galactosidase of Sulfolobus solfataricus is only one of several activities of a thermostable β-D-glycodiase. Appl. Environ. Microbiol. 57, 1644-1649). Members of the latter group, although highly specific with respect to the β-anomeric configuration of the glycosidic linkage, often display a rather relaxed substrate specificity and hydrolyze βglucosides as well as β -fucosides and β -galactosides.

10

15

20

Generally, α -galactosidases are enzymes that catalyze the hydrolysis of galactose groups on a polysaccharide backbone or hydrolyze the cleavage of di- or oligosaccharides comprising galactose.

Generally, \(\beta\)-mannanases are enzymes that catalyze the hydrolysis of mannose groups internally on a polysaccharide backbone or hydrolyze the cleavage of di- or oligosaccaharides comprising mannose groups. \(\beta\)-mannosidases hydrolyze non-reducing, terminal mannose residues on a mannose-containing polysaccharide and the cleavage of di- or oligosaccaharides comprising mannose groups.

Guar gum is a branched galactomannan polysaccharide composed of β -1,4 linked mannose backbone with α -1,6 linked galactose side chains. The enzymes required for the degradation of guar are β -mannanase, β -mannosidase and α -galactosidase. β -mannanase hydrolyses the mannose backbone internally and β -mannosidase hydrolyses non-reducing, terminal mannose residues. α -galactosidase hydrolyses α -linked galactose groups.

Galactomannan polysaccharides and the enzymes that degrade them have a variety of applications. Guar is commonly used as a thickening agent in food and is utilized in hydraulic fracturing in oil and gas recovery. Consequently, galactomannanases are industrially relevant for the degradation and modification of guar. Furthermore, a need exists for thermostable galactomannases that are active in extreme conditions associated with drilling and well stimulation.

There are other applications for these enzymes in various industries, such as in the beet sugar industry. 20-30% of the domestic U.S. sucrose consumption is sucrose from sugar beets. Raw beet sugar can contain a small amount of raffinose when the sugar beets are stored before processing and rotting begins to set in. Raffinose inhibits the crystallization of sucrose and also constitutes a hidden quantity of sucrose. Thus, there is merit to eliminating raffinose from raw beet sugar. α -Galactosidase has also been used as a digestive aid to break down raffinose, stachyose, and verbascose in such foods as beans and other gassy foods.

 β -galactosidases which are active and stable at high temperatures appear to be superior enzymes for the production of lactose-free dietary milk products (Chaplin, M.F.

10

15

20

and Bucke, C. (1990) In: Enzyme Technology, pp. 159-160, Cambridge University Press, Cambridge, UK). Also, several studies have demonstrated the applicability of β-galactosidases to the enzymatic synthesis of oligosaccharides via transglycosylation reactions (Nilsson, K.G.I. (1988) Enzymatic synthesis of oligosaccharides. Trends Biotechnol. 6, 156-264; Cote, G.L. and Tao, B.Y. (1990) Oligosaccharide synthesis by enzymatic transglycosylation. Glycoconjugate J. 7, 145-162). Despite the commercial potential, only a few β-galactosidases of thermophiles have been characterized so far. Two genes reported are β-galactoside-cleaving enzymes of the hyperthermophilic bacterium *Thermotoga maritima*, one of the most thermophilic organotrophic eubacteria described to date (Huber, R., Langworthy, T.A., König, H., Thomm, M., Woese, C.R., Sleytr, U.B. and Stetter, K.O. (1986) *T. martima* sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90°C, Arch. Microbiol. 144, 324-333) one of the most thermophilic organotrophic eubacteria described to date. The gene products have been identified as a β-galactosidase and a β-glucosidase.

Pullulanase is well known as a debranching enzyme of pullulan and starch. The enzyme hydrolyzes α -1,6-glucosidic linkages on these polymers. Starch degradation for the production or sweeteners (glucose or maltose) is a very important industrial application of this enzyme. The degradation of starch is developed in two stages. The first stage involves the liquefaction of the substrate with α -amylase, and the second stage, or saccharification stage, is performed by β -amylase with pullalanase added as a debranching enzyme, to obtain better yields.

Endoglucanases can be used in a variety of industrial applications. For instance, the endoglucanases of the present invention can hydrolyze the internal ß-1,4-glycosidic bonds in cellulose, which may be used for the conversion of plant biomass into fuels and chemicals. Endoglucanases also have applications in detergent formulations, the textile industry, in animal feed, in waste treatment, and in the fruit juice and brewing industry for the clarification and extraction of juices.

10

15

20

5

10

15

20

25

Brief Description of the Drawings

The following drawings are illustrative of embodiments of the invention and are not meant to limit the scope of the invention as encompassed by the claims.

Figures 1a-b are the full-length DNA and corresponding deduced amino acid sequence of M11TL of the present invention. Sequencing was performed using a 378 automated DNA sequencer for all sequences of the present invention (Applied Biosystems, Inc.).

Figure 2 is an illustration of the full-length DNA and corresponding deduced amino acid sequence of OC1/4V-33B/G.

Figure 3 is an illustration of the full-length DNA and corresponding deduced amino acid sequence of F1-12G.

Figures 4a-b are the full-length DNA and corresponding deduced amino acid sequence of 9N2-31B/G.

Figures 5a-b are the full-length DNA and corresponding deduced amino acid sequence of MSB8-6G.

Figure 6 is the full-length DNA and corresponding deduced amino acid sequence of AEDII12RA-18B/G.

Figures 7a-b are the full-length DNA and corresponding deduced amino acid sequence of GC74-22G.

Figures 8a-b are the full-length DNA and corresponding deduced amino acid sequence of VC1-7G1.

Figures 9a-c are the full-length DNA and corresponding deduced amino acid sequence of 37GP1.

Figures 10a-c are the full-length DNA and corresponding deduced amino acid sequence of 6GC2.

Figures 11a-d are the full-length DNA and corresponding deduced amino acid sequence of 6GP2.

Figures 12a-c are the full-length DNA and corresponding deduced amino acid sequence of 63GB1.

Figures 13a-b are the full-length DNA and corresponding deduced amino acid sequence of OC1/4V.

Figures 14a-e are the full-length DNA and corresponding deduced amino acid sequence of 6GP3.

Figures 15a-d are the full-length DNA and corresponding deduced amino acid sequence of *Thermotoga maritima* MSB8-6GP2.

Figures 16a-c are the full-length DNA and corresponding deduced amino acid sequence of *Thermotoga maritima* MSB8-6GB4.

Figures 17a-d are the full-length DNA and corresponding deduced amino acid sequence of *Banki gouldi* 37GP4.

Figures 18a-b are the full-length DNA and corresponding deduced amino acid sequence of *Pyrococcus furiosus* VC1-7EG1.

SUMMARY OF THE INVENTION

In a preferred embodiment of the present invention, there are provided isolated nucleic acids (polynucleotides) which encode mature enzymes having the deduced amino acid sequences of Figures 1-18 (SEQ ID NOS: 15-28 and 61-64).

In another embodiment, the invention provides a method for producing a polypeptide including culturing host cells containing the polynucleotide of Figures 1-18 and expressing from the host cell a polypeptide encoded by the polynucleotide and isolating the polypeptide.

In another embodiment, the invention provides an enzyme selected from the group consisting of an enzyme having an amino acid sequence set forth in SEQ ID NOS: 15-28 or 61-64 and an enzyme which has at least 30 consecutive amino acid residue as an enzyme having an amino acid sequence set forth in SEQ ID NOS: 15-28 or 61-64.

In yet another embodiment, the invention provides a method for generating glucose from soluble cell oligosaccharides which includes contacting a sample containing oligosaccharides with an effective amount of an enzyme selected from the group of

5.

10

15

20

enzymes having the amino acid sequence set forth in SEQ ID NOS: 15-28, 61-63 and 64 such that glucose is produced

The publications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

Definitions

"Monosaccharide", as used herein, refers to a single polyhydroxy aldehyde or ketone unit.

"Oligosaccharide", as used herein, consist of short chains of monosaccharide units joined together by covalent bonds. Of these, the most abundant are the disaccharides, which have two monosaccharide units.

"Polysaccharide", as used herein, consists of long chains having many monosaccharide units.

The term "gene" means the segment of DNA involved in producing a polypeptide chain; it includes regions preceding and following the coding region (leader and trailer) as well as intervening sequences (introns) between individual coding segments (exons).

A coding sequence is "operably linked to" another coding sequence when RNA polymerase will transcribe the two coding sequences into a single mRNA, which is then translated into a single polypeptide having amino acids derived from both coding sequences. The coding sequences need not be contiguous to one another so long as the expressed sequences ultimately process to produce the desired protein.

"Recombinant" enzymes refer to enzymes produced by recombinant DNA techniques; *i.e.*, produced from cells transformed by an exogenous DNA construct encoding the desired enzyme. "Synthetic" enzymes are those prepared by chemical synthesis.

A DNA "coding sequence of" or a "nucleotide sequence encoding" a particular enzyme, is a DNA sequence which is transcribed and translated into an enzyme when placed under the control of appropriate regulatory sequences.

15

20

Detailed Description of the Invention

The polynucleotides and polypeptides of the present invention have been identified as glucosidases. α -galactosidases. β -galactosidases, β -mannosidases, β -mannanases, endoglucanases, and pullalanases as a result of their enzymatic activity.

In accordance with one aspect of the present invention, there are provided novel enzymes, as well as active fragments, analogs and derivatives thereof.

In accordance with another aspect of the present invention, there are provided isolated nucleic acid molecules encoding the enzymes of the present invention including mRNAs, cDNAs, genomic DNAs as well as active analogs and fragments of such enzymes.

In accordance with yet a further aspect of the present invention, there is provided a process for producing such polypeptides by recombinant techniques comprising culturing recombinant prokaryotic and/or eukaryotic host cells, containing a nucleic acid sequence of the present invention, under conditions promoting expression of said enzymes and subsequent recovery of said enzymes.

In accordance with yet a further aspect of the present invention, there is provided a process for utilizing such enzymes, or polynucleotides encoding such enzymes for hydrolyzing lactose to galactose and glucose for use in the food processing industry, the pharmaceutical industry, for example, to treat intolerance to lactose, as a diagnostic reporter molecule, in com wet milling, in the fruit juice industry, in baking, in the textile industry and in the detergent industry.

In accordance with yet a further aspect of the present invention, there is provided a process for utilizing such enzymes for hydrolyzing guar gum (a galactomannan polysaccharide) to remove non-reducing terminal mannose residues. Further polysaccharides such as galactomannan and the enzymes according to the invention that degrade them have a variety of applications. Guar gum is commonly used as a thickening agent in food and also is utilized in hydraulic fracturing in oil and gas recovery. Consequently, mannanases are industrially relevant for the degradation and modification of guar gums. Furthermore, a need exists for thermostable mannases that are active in extreme conditions associated with drilling and well stimulation.

10

15

20

In accordance with yet a further aspect of the present invention, there are also provided nucleic acid probes comprising nucleic acid molecules of sufficient length to specifically hybridize to a nucleic acid sequence of the present invention.

In accordance with yet a further aspect of the present invention, there is provided a process for utilizing such enzymes, or polynucleotides encoding such enzymes, for *in vitro* purposes related to scientific research. for example, to generate probes for identifying similar sequences which might encode similar enzymes from other organisms by using certain regions, *i.e.*, conserved sequence regions, of the nucleotide sequence.

These and other aspects of the present invention should be apparent to those skilled in the art from the teachings herein.

The polynucleotides of this invention were originally recovered from genomic gene libraries derived from the following organisms:

M11TL is a new species of *Desulfurococcus* isolated from Diamond Pool in Yellowstone National Park. The organism grows optimally at 85-88°C, pH 7.0 in a low salt medium containing yeast extract, peptone, and gelatin as substrates with a N₂/CO₂ gas phase.

OC1/4V is from the genus *Thermotoga*. The organism was isolated from Yellowstone National Park. It grows optimally at 75°C in a low salt medium with cellulose as a substrate and N_2 in gas phase.

Pyrococcus furiosus VC1 and (7EG1) is from the genus Pyrococcus. VC1 was isolated from Vulcano, Italy. It grows optimally at 100°C in a high salt medium (marine) containing elemental sulfur, yeast extract, peptone and starch as substrates and N₂ in gas phase.

Staphylothermus marinus F1 is a from the genus Staphylothermus. F1 was isolated from Vulcano, Italy. It grows optimally at 85°C, pH 6.5 in high salt medium (marine) containing elemental sulfur and yeast extract as substrates and N₂ in gas phase.

10

15

20

Thermococcus 9N-2 is from the genus Thermococcus 9N-2 was isolated from diffuse vent fluid in the East Pacific Rise. It is a strict anaerobe that grows optimally at 87°C.

Thermotoga maritima MSB8 and MSB8 (Clone # 6GP2 and 6GB4) is from the genus Thermotogo, and was isolated from Vulcano, Italy. MSB8 grows optimally at 85°C, pH 6.5 in a high salt medium (marine) containing starch and yeast extract as substrates and N, in gas phase.

Thermococcus alcaliphilus AEDII12RA is from the genus Thermococcus. AEDII12RA grows optimally at 85° C, pH 9.5 in a high salt medium (marine) containing polysulfides and yeast extract as substrates and N_2 in gas phase.

Thermococcus chitonophagus GC74 is from the genus Thermococcus. GC74 grows optimally at 85°C, pH 6.0 in a high salt medium (marine) containing chitin, meat extract, elemental sulfur and yeast extract as substrates and N₂ in gas phase. AEPII 1a grows optimally at 85°C at pH 6.5 in marine medium under anaerobic conditions. It has many substrates. Bankia gouldi is from the genus Bankia.

Accordingly, the polynucleotides and enzymes encoded thereby are identified by the organism from which they were isolated, and are sometimes hereinafter referred to as "M11TL" (Figure 1 and SEQ ID NOS:1 and 15), "OC1/4V-33B/G" (Figure 2 and SEQ ID NOS:2 and 16), "F1-12G" (Figure 3 and SEQ ID NOS:3 and 17), "9N2-31B/G" (Figure 4 and SEQ ID NOS:4 and 18), "MSB8" (Figure 5 and SEQ ID NOS:5 and 19), "AEDII12RA-18B/G" (Figure 6 and SEQ ID NOS:6 and 20), "GC74-22G" (Figure 7 and SEQ ID NOS:7 and 21), "VC1-7G1" (Figure 8 and SEQ ID NOS:8 and 22), "37GP1" (Figure 9 and SEQ ID NOS: 9 and 23), "6GC2" (Figure 10 and SEQ ID NOS: 10 and 24), "6GP2" (Figure 11 and SEQ ID NOS:11 and 25), "AEPII 1a" (Figure 12 and SEQ ID NOS:12 and 26), "OC1/4V" (Figure 13 and SEQ ID NOS:13 and 27), and "6GP3" (Figure 14 and SEQ ID NOS:28), "MSB8-6GP2" (Figure 15 and SEQ ID NOS:57 and 61), "MSB8-6GB4"(Figure 16 and SEQ ID NOS:58 and 62),"VC1-7EG1"(Figure 17 and SEQ ID NOS:59 and 63), and 37GP4 (Figure 18 and SEQ ID NOS:60 and 64).

5

10

15

20

The polynucleotides and polypeptides of the present invention show identity at the nucleotide and protein level to known genes and proteins encoded thereby as shown in Table 1.

Table 1

	T		
	Gene/Protein with	Protein	Nucleic Acid
Clone	Closest Homology	Identity	Identity
M11TL-29G	Sulfolobus sulfataricus	51%	55%
	DSM 1616/P1, β-		·
	galactosidase		
OC1/4V-33B/G	Caldocellum	52%	57%
	saccharolyticum, β-		
	glucosidase		
Staphylothermus	Bacillus polymyxa, β-	36%	48%
marinus F1-12G	galactosidase		
Thermococcus 9N2-	Sulfolobus sulfataricus	51%	50%
31B/G	ATCC 49255/MT4, β-	·	
	galactosidase		
Thermotoga maritima	Clostridium thermocellum	45%	53%
MSB8-6G	bglB		
Thermococcus	Bacillus polymyxa, β-	34%	48%
AEDII12RA-18B/G	galactosidase		
Thermococcus	Sulfolobus sulfataricus	46%	54%
chitonophagus GC74-	ATCC 49255/MT4, β-		
22G	galactosidase		

15

Pyrococcus furiosus VC1-7G1	Sulfolobus sulfataricus/MT-4 β- galactosidase	46.4%	52.5%
Thermotoga maritima α-galactosidase (6GC2)	Pediococcus pentosaceaus α-galactosidase	49%	29%
Thermotoga maritima ß-mannanase (6GP2)	Aspergillus aculeatus mannanase	56%	37%
AEPII 1a ß- mannosidase (63GB1)	Sulfolobus solfactaricus ß- galactosidase	78%	56%
OC1/4V endoglucanase (33GP1)	Clostridium thermocellum endo-1,4-ß-endoglucanase	65%	43%
Thermotoga maritima pullalanase (6GP3)	Caldocellum saccharolyticum α- destrom 6 glucanohydralase	72	53
Bankia gouldi mix Endoglucanase (37GP1)	None available		

The polynucleotides and enzymes of the present invention show homology to each other as shown in Table 2.

10

Table 2

Clone	Gene/Protein with Closest Homology	Protein Identity	Nucleic Acid Identity
Staphylothermus marinus F1-12G	Thermococcus AEDII12RA-18B/G, β- galactosidase, glucosidase	55%	57%
Thermococcus 9N2- 31B/G	Thermococcus chitonophagus GC74- 22G-glucosidase`	74%	66%
Pyrococcus furiosus VC1-7G1	Pyrococcus furiosus VC1- 7B/G β-galactosidase	46.4%	54%

All the clones identified in Tables 1 and 2 encode polypeptides which have α -glycosidase or β -glycosidase activity.

This invention, in addition to the isolated nucleic acid molecules encoding the enzymes of the present invention, also provide substantially similar sequences. Isolated nucleic acid sequences are substantially similar if: (i) they are capable of hybridizing under conditions hereinafter described, to the polynucleotides of SEQ ID NOS: 1-14 and 57-60; (ii) or they encode DNA sequences which are degenerate to the polynucleotides of SEQ ID NOS: 1-14 and 57-60. Degenerate DNA sequences encode the amino acid sequences of SEQ ID NOS:15-28 and 61-64, but have variations in the nucleotide coding sequences. As used herein, substantially similar refers to the sequences having similar identity to the sequences of the instant invention. The nucleotide sequences that are substantially the same can be identified by hybridization or by sequence comparison. Enzyme sequences that are substantially the same can be identified by one or more of the following: proteolytic digestion, gel electrophoresis and/or microsequencing.

One means for isolating the nucleic acid molecules encoding the enzymes of the present invention is to probe a gene library with a natural or artificially designed probe using art recognized procedures (see, for example: Current Protocols in Molecular Biology,

15

20

Ausubel F.M. et al. (EDS.) Green Publishing Company Assoc. and John Wiley Interscience, New York, 1989, 1992). It is appreciated to one skilled in the art that the polynucleotides of SEQ ID NOS: 1-14 and 57-60 or fragments thereof (comprising at least 12 contiguous nucleotides), are particularly useful probes. Other particular useful probes for this purpose are hybridizable fragments to the sequences of SEQ ID NOS: 1-14 and 57-60 (i.e., comprising at least 12 contiguous nucleotides).

With respect to nucleic acid sequences which hybridize to specific nucleic acid sequences disclosed herein, hybridization may be carried out under conditions of reduced stringency, medium stringency or even stringent conditions. As an example of oligonucleotide hybridization, a polymer membrane containing immobilized denatured nucleic acids is first prehybridized for 30 minutes at 45°C in a solution consisting of 0.9 M NaCl, 50 mM NaH₂PO₄, pH 7.0, 5.0 mM Na₂EDTA, 0.5% SDS, 10X Denhardt's, and 0.5 mg/ml polyriboadenylic acid. Approximately 2 X 10⁷ cpm (specific activity 4-9 X 10 cpm/ug) of ³²P end-labeled oligonucleotide probe are then added to the solution. After 12-16 hours of incubation, the membrane is washed for 30 minutes at room temperature in 1X SET (150 mM NaCl, 20 mM Tris hydrochloride, pH 7.8, 1 mM Na₂EDTA) containing 0.5% SDS, followed by a 30 minute wash in fresh 1X SET at Tm 10°C for the oligonucleotide probe. The membrane is then exposed to auto-radiographic film for detection of hybridization signals.

Stringent conditions means hybridization will occur only if there is at least 90% identity, preferably at least 95% identity and most preferably at least 97% identity between the sequences. Further, it is understood that a section of a 100 bps sequence that is 95 bps in length has 95% identity with the 1090 bps sequence from which it is obtained. See J. Sambrook et al., Molecular Cloning, A Laboratory Manual, 2d Ed., Cold Spring Harbor Laboratory (1989) which is hereby incorporated by reference in its entirety. Also, it is understood that a fragment of a 100 bps sequence that is 95 bps in length has 95% identity with the 100 bps sequence from which it is obtained.

As used herein, a first DNA (RNA) sequence is at least 70% and preferably at least 80% identical to another DNA (RNA) sequence if there is at least 70% and preferably at

5

10

15

20

least a 80% or 90% identity, respectively, between the bases of the first sequence and the bases of the another sequence, when properly aligned with each other, for example when aligned by BLASTN.

"Identity" as the term is used herein, refers to a polynucleotide sequence which comprises a percentage of the same bases as a reference polynucleotide (SEQ ID NOS:1-14 and 57-60). For example, a polynucleotide which is at least 90% identical to a reference polynucleotide, has polynucleotide bases which are identical in 90% of the bases which make up the reference polynucleotide and may have different bases in 10% of the bases which comprise that polynucleotide sequence.

The present invention relates polynucleotides which differ from the reference polynucleotide such that the changes are silent changes, for example the change do not alter the amino acid sequence encoded by the polynucleotide. The present invention also relates to nucleotide changes which result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference polynucleotide. In a preferred aspect of the invention these polypeptides retain the same biological action as the polypeptide encoded by the reference polynucleotide.

It is also appreciated that such probes can be and are preferably labeled with an analytically detectable reagent to facilitate identification of the probe. Useful reagents include but are not limited to radioactivity, fluorescent dyes or enzymes capable of catalyzing the formation of a detectable product. The probes are thus useful to isolate complementary copies of DNA from other sources or to screen such sources for related sequences.

The polynucleotides of this invention were recovered from genomic gene libraries from the organisms listed in Table 1. For example, gene libraries can be generated in the Lambda ZAP II cloning vector (Stratagene Cloning Systems). Mass excisions can be performed on these libraries to generate libraries in the pBluescript phagemid. Libraries are thus generated and excisions performed according to the protocols/methods hereinafter described.

5

10

15

20

The excision libraries are introduced into the *E. coli* strain BW14893 F'kan1A. Expression clones are then identified using a high temperature filter assay. Expression clones encoding several glucanases and several other glycosidases are identified and repurified. The polynucleotides, and enzymes encoded thereby, of the present invention, yield the activities as described above.

The coding sequences for the enzymes of the present invention were identified by screening the genomic DNAs prepared for the clones having glucosidase or galactosidase activity.

An example of such an assay is a high temperature filter assay wherein expression clones were identified by use of high temperature filter assays using buffer Z (see recipe below) containing 1 mg/ml of the substrate 5-bromo-4-chloro-3-indolyl-β-D-glucopyranoside (XGLU) (Diagnostic Chemicals Limited or Sigma) after introducing an excision library into the *E. coli* strain BW14893 F'kan1A. Expression clones encoding XGLUases were identified and repurified from M11TL, OC1/4V, Pyrococcus furiosus VC1, Staphylothemus marinus F1, Thermococcus 9N-2, Thermotoga maritima MSB8, Thermococcus alcaliphilus AEDII12RA, and Thermococcus chitonophagus GC74.

Z-buffer: (referenced in Miller, J.H. (1992) A Short Course in Bacterial Genetics, p. 445.)

per liter:

Adjust pH to 7.0

10

15

20

25

 $Na_2HPO_4-7H_2O$ 16.1g $NaH_2PO_4-7H_2O$ 5.5g KCl 0.75g MgSO₄-7H₂O 0.246g β-mercaptoethanol 2.7ml

High Temperature Filter Assay

(1) The f factor fkan (from E. coli strain CSH118)(1) was introduced into the pho-pnh-lac-strain BW14893(2). BW13893(2). The filamentous phage library was plated on the resulting strain, BW14893 F'kan. (Miller, J.H. (1992) A Short Course in

Bacterial Genetics; Lee, K.S., Metcalf, et al., (1992) Evidence for two phosphonate degradative pathways in Enterobacter Aerogenes, J. Bacteriol., 174:2501-2510.

- (2) After growth on 100 mm LB plates containing 100 μg/ml ampicillin, 80 μg/ml nethicillin and 1mM IPTG, colony lifts were performed using Millipore HATF membrane filters.
- (3) The colonies transferred to the filters were lysed with chloroform vapor in 150 mm glass petri dishes.
- (4) The filters were transferred to 100 mm glass petri dishes containing a piece of Whatman 3MM filter paper saturated with buffer.
 - (a) when testing for galactosidase activity (XGALase), 3MM paper was saturated with Z buffer containing 1 mg/ml XGAL (ChemBridge Corporation). After transferring filter bearing lysed colonies to the glass petri dish, placed dish in oven at 80-85°C.
 - (b) when testing for glucosidase (XGLUase), 3MM paper was saturated with Z buffer containing 1 mg/ml XGLU. After transferring filter bearing lysed colonies to the glass petri dish, placed dish in oven at 80-85°C.
- (5) Positives' were observed as blue spots on the filter membranes. Used the following filter rescue technique to retrieve plasmid from lysed positive colony. Used pasteur pipette (or glass capillary tube) to core blue spots on the filter membrane. Placed the small filter disk in an Eppendorf tube containing 20 μl water. Incubated the Eppendorf tube at 75°C for 5 minutes followed by vortexing to elute plasmid DNA off filter. This DNA was transformed into electrocompetent *E. coli* cells DH10B for Thermatoga maritima MSB8-6G, Staphylothermus marinus F1-12G, Thermococcus AEDII12RA-18B/G, Thermococcus chitonophagus GC74-22G, M11Tl and OC1/4V. Electrocompetent BW14893 F'kan1A *E. coli* were used for Thermococcus 9N2-31B/G, and *Pyrococcus furiosus* VC1-7G1. Repeated filter-lift assay on transformation plates to identify 'positives'. Return transformation plates to 37°C incubator after filter lift to regenerate colonies. Inoculate 3 ml LB liquid containing 100 μg/ml ampicillin with repurified positives and incubate at 37°C

5

10

15

20

overnight. Isolate plasmid DNA from these cultures and sequence plasmid insert. In some instances where the plates used for the initial colony lifts contained non-confluent colonies, a specific colony corresponding to a blue spot on the filter could be identified on a regenerated plate and repurified directly, instead of using the filter rescue technique.

Another example of such an assay is a variation of the high temperature filter assay wherein colony-laden filters are heat-killed at different temperatures (for example, 105°C for 20 minutes) to monitor thermostability. The 3MM paper is saturated with different buffers (i.e., 100 mM NaCl, 5 mM MgCl₂, 100 mM Tris-Cl (pH 9.5)) to determine enzyme activity under different buffer conditions.

10

15

20

25

A β -glucosidase assay may also be employed, wherein Glcp β Np is used as an artificial substrate (aryl- β -glucosidase). The increase in absorbance at 405 nm as a result of p-nitrophenol (pNp) liberation was followed on a Hitachi U-1100 spectrophotometer, equipped with a thermostatted cuvette holder. The assays may be performed at 80°C or 90°C in closed 1-ml quartz cuvette. A standard reaction mixture contains 150 mM trisodium substrate, pH 5.0 (at 80°C), and 0.95 mM pNp derivative pNp = 0.561 mM⁻¹ cm⁻¹). The reaction mixture is allowed to reach the desired temperature, after which the reaction is started by injecting an appropriate amount of enzyme (1.06 ml final volume).

1 U β -glucosidase activity is defined as that amount required to catalyze the formation of 1.0 μ mol pNp/min. D-cellobiose may also be used as a substrate.

An ONPG assay for β -galactosidase activity is described by Miller, J.H. (1992) A Short Course in Bacterial Genetics and Mill, J.H. (1992) Experiments in Molecular Genetics, the contents of which are hereby incorporated by reference in their entirety.

A quantitative fluorometric assay for β -galactosidase specific activity is described by : Youngman P., (1987) Plasmid Vectors for Recovering and Exploiting Tn917 Transpositions in Bacillus and other Gram-Positive Bacteria. In Plasmids: A Practical approach (ed. K. Hardy) pp 79-103. IRL Press, Oxford. A description of the procedure can be found in Miller (1992) p. 75-77, the contents of which are incorporated by reference herein in their entirety.

The polynucleotides of the present invention may be in the form of DNA which DNA includes cDNA, genomic DNA, and synthetic DNA. The DNA may be double-stranded or single-stranded, and if single stranded may be the coding strand or non-coding (anti-sense) strand. The coding sequences which encodes the mature enzymes may be identical to the coding sequences shown in Figures 1-8 (SEQ ID NOS: 1-14 and 57-60) or may be a different coding sequence which coding sequence, as a result of the redundancy or degeneracy of the genetic code, encodes the same mature enzymes as the DNA of Figures 1-18 (SEQ ID NOS: 1-14 and 57-60).

The polynucleotide which encodes for the mature enzyme of Figures 1-18 (SEQ ID NOS: 15-28 and 61-64) may include, but is not limited to: only the coding sequence for the mature enzyme; the coding sequence for the mature enzyme and additional coding sequence such as a leader sequence or a proprotein sequence; the coding sequence for the mature enzyme (and optionally additional coding sequence) and non-coding sequence, such as introns or non-coding sequence 5' and/or 3' of the coding sequence for the mature enzyme.

Thus, the term "polynucleotide encoding an enzyme (protein)" encompasses a polynucleotide which includes only coding sequence for the enzyme as well as a polynucleotide which includes additional coding and/or non-coding sequence.

The present invention further relates to variants of the hereinabove described polynucleotides which encode for fragments, analogs and derivatives of the enzymes having the deduced amino acid sequences of Figures 1-18 (SEQ ID NOS: 15-28 and 61-64). The variant of the polynucleotide may be a naturally occurring allelic variant of the polynucleotide or a non-naturally occurring variant of the polynucleotide.

Thus, the present invention includes polynucleotides encoding the same mature enzymes as shown in Figures 1-18 (SEQ ID NOS: 15-28 and 61-64) as well as variants of such polynucleotides which variants encode for a fragment, derivative or analog of the enzymes of Figures 1-18 (SEQ ID NOS: 15-28 and 61-64). Such nucleotide variants include deletion variants, substitution variants and addition or insertion variants.

As hereinabove indicated, the polynucleotides may have a coding sequence which is a naturally occurring allelic variant of the coding sequences shown in Figures 1-18 (SEQ

5

10

15

20

ID NOS: 1-14 and 57-60). As known in the art, an allelic variant is an alternate form of a polynucleotide sequence which may have a substitution, deletion or addition of one or more nucleotides, which does not substantially alter the function of the encoded enzyme.

. 5

15

20

25

Fragments of the full length gene of the present invention may be used as a hybridization probe for a cDNA or a genomic library to isolate the full length DNA and to isolate other DNAs which have a high sequence similarity to the gene or similar biological activity. Probes of this type preferably have at least 10, preferably at least 15, and even more preferably at least 30 bases and may contain, for example, at least 50 or more bases. The probe may also be used to identify a DNA clone corresponding to a full length transcript and a genomic clone or clones that contain the complete gene including regulatory and promotor regions, exons, and introns. An example of a screen comprises isolating the coding region of the gene by using the known DNA sequence to synthesize an oligonucleotide probe. Labeled oligonucleotides having a sequence complementary to that of the gene of the present invention are used to screen a library of genomic DNA to determine which members of the library the probe hybridizes to.

The present invention further relates to polynucleotides which hybridize to the hereinabove-described sequences if there is at least 70%, preferably at least 90%, and more preferably at least 95% identity between the sequences. The present invention particularly relates to polynucleotides which hybridize under stringent conditions to the hereinabove-described polynucleotides. As herein used, the term "stringent conditions" means hybridization will occur only if there is at least 95% and preferably at least 97% identity between the sequences. The polynucleotides which hybridize to the hereinabove described polynucleotides in a preferred embodiment encode enzymes which either retain substantially the same biological function or activity as the mature enzyme encoded by the DNA of Figures 1-18 (SEQ ID NOS: 1-14 and 57-60).

Alternatively, the polynucleotide may have at least 15 bases, preferably at least 30 bases, and more preferably at least 50 bases which hybridize to any part of a polynucleotide of the present invention and which has an identity thereto, as hereinabove described, and which may or may not retain activity. For example, such polynucleotides may be employed

as probes for the polynucleotides of SEQ ID NOS: 1-14 and 57-60, for example, for recovery of the polynucleotide or as a diagnostic probe or as a PCR primer.

Thus, the present invention is directed to polynucleotides having at least a 70% identity, preferably at least 90% identity and more preferably at least a 95% identity to a polynucleotide which encodes the enzymes of SEQ ID NOS: 15-28 and 61-64 as well as fragments thereof, which fragments have at least 15 bases, preferably at least 30 bases and most preferably at least 50 bases, which fragments are at least 90% identical, preferably at least 95% identical and most preferably at least 97% identical under stringent conditions to any portion of a polynucleotide of the present invention.

The present invention further relates to enzymes which have the deduced amino acid sequences of Figures 1-18 (SEQ ID NOS: 15-28 and 61-64) as well as fragments, analogs and derivatives of such enzyme.

The terms "fragment," "derivative" and "analog" when referring to the enzymes of Figures 1-18 (SEQ ID NOS: 15-28 and 61-64) means enzymes which retain essentially the same biological function or activity as such enzymes. Thus, an analog includes a proprotein which can be activated by cleavage of the proprotein portion to produce an active mature enzyme.

The enzymes of the present invention may be a recombinant enzyme, a natural enzyme or a synthetic enzyme, preferably a recombinant enzyme.

The fragment, derivative or analog of the enzymes of Figures 1-18 (SEQ ID NOS: 15-28 and 61-64) may be (i) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code, or (ii) one in which one or more of the amino acid residues includes a substituent group, or (iii) one in which the mature enzyme is fused with another compound, such as a compound to increase the half-life of the enzyme (for example, polyethylene glycol), or (iv) one in which the additional amino acids are fused to the mature enzyme, such as a leader or secretory sequence or a sequence which is employed for purification of the mature enzyme or a proprotein sequence. Such fragments, derivatives

5

10

15

20

and analogs are deemed to be within the scope of those skilled in the art from the teachings herein.

The enzymes and polynucleotides of the present invention are preferably provided in an isolated form, and preferably are purified to homogeneity.

The term "isolated" means that the material is removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide or enzyme present in a living animal is not isolated, but the same polynucleotide or enzyme, separated from some or all of the coexisting materials in the natural system, is isolated. Such polynucleotides could be part of a vector and/or such polynucleotides or enzymes could be part of a composition, and still be isolated in that such vector or composition is not part of its natural environment.

The enzymes of the present invention include the enzymes of SEQ ID NOS: 15-28 and 61-64 (in particular the mature enzyme) as well as enzymes which have at least 70% similarity (preferably at least 70% identity) to the enzymes of SEQ ID NOS: 15-28 and 61-64 and more preferably at least 90% similarity (more preferably at least 90% identity) to the enzymes of SEQ ID NOS: 15-28 and 61-64 and still more preferably at least 95% similarity (still more preferably at least 95% identity) to the enzymes of SEQ ID NOS: 15-28 and 61-64 and also include portions of such enzymes with such portion of the enzyme generally containing at least 30 amino acids and more preferably at least 50 amino acids.

As known in the art "similarity" between two enzymes is determined by comparing the amino acid sequence and its conserved amino acid substitutes of one enzyme to the sequence of a second enzyme.

A variant, i.e. a "fragment", "analog" or "derivative" polypeptide, and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions, fusions and truncations, which may be present in any combination.

Among preferred variants are those that vary from a reference by conservative amino acid substitutions. Such substitutions are those that substitute a given amino acid in a polypeptide by another amino acid of like characteristics. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala,

5

10

15

20

Val, Leu and Ile; interchange of the hydroxyl residues Ser and Thr, exchange of the acidic residues Asp and Glu, substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg and replacements among the aromatic residues Phe, Tyr.

Most highly preferred are variants which retain the same biological function and activity as the reference polypeptide from which it varies.

Fragments or portions of the enzymes of the present invention may be employed for producing the corresponding full-length enzyme by peptide synthesis: therefore, the fragments may be employed as intermediates for producing the full-length enzymes. Fragments or portions of the polynucleotides of the present invention may be used to synthesize full-length polynucleotides of the present invention.

The present invention also relates to vectors which include polynucleotides of the present invention, host cells which are genetically engineered with vectors of the invention and the production of enzymes of the invention by recombinant techniques.

Host cells are genetically engineered (transduced or transformed or transfected) with the vectors of this invention which may be, for example, a cloning vector or an expression vector. The vector may be, for example, in the form of a plasmid, a viral particle, a phage, etc. The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants or amplifying the genes of the present invention. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to the ordinarily skilled artisan.

The polynucleotides of the present invention may be employed for producing enzymes by recombinant techniques. Thus, for example, the polynucleotide may be included in any one of a variety of expression vectors for expressing an enzyme. Such vectors include chromosomal, nonchromosomal and synthetic DNA sequences, e.g., derivatives of SV40; bacterial plasmids; phage DNA; baculovirus; yeast plasmids; vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia, adenovirus, fowl pox virus, and pseudorabies. However, any other vector may be used as long as it is replicable and viable in the host.

5

10

15

20

The appropriate DNA sequence may be inserted into the vector by a variety of procedures. In general, the DNA sequence is inserted into an appropriate restriction endonuclease site(s) by procedures known in the art. Such procedures and others are deemed to be within the scope of those skilled in the art.

The DNA sequence in the expression vector is operatively linked to an appropriate expression control sequence(s) (promoter) to direct mRNA synthesis. As representative examples of such promoters, there may be mentioned: LTR or SV40 promoter, the <u>E. colilac</u> or <u>trp</u>, the phage lambda P_L promoter and other promoters known to control expression of genes in prokaryotic or eukaryotic cells or their viruses. The expression vector also contains a ribosome binding site for translation initiation and a transcription terminator. The vector may also include appropriate sequences for amplifying expression.

10

15

20

25

In addition, the expression vectors preferably contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells such as dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, or such as tetracycline or ampicillin resistance in <u>E. coli</u>.

The vector containing the appropriate DNA sequence as hereinabove described, as well as an appropriate promoter or control sequence, may be employed to transform an appropriate host to permit the host to express the protein.

As representative examples of appropriate hosts, there may be mentioned: bacterial cells, such as <u>E. coli</u>, <u>Streptomyces</u>, <u>Bacillus subtilis</u>; fungal cells, such as yeast: insect cells such as <u>Drosophila S2</u> and <u>Spodoptera Sf9</u>; animal cells such as CHO, COS or Bowes melanoma; adenoviruses; plant cells, etc. The selection of an appropriate host is deemed to be within the scope of those skilled in the art from the teachings herein.

More particularly, the present invention also includes recombinant constructs comprising one or more of the sequences as broadly described above. The constructs comprise a vector, such as a plasmid or viral vector, into which a sequence of the invention has been inserted, in a forward or reverse orientation. In a preferred aspect of this embodiment, the construct further comprises regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and

promoters are known to those of skill in the art, and are commercially available. The following vectors are provided by way of example; Bacterial: pQE70, pQE60, pQE-9 (Qiagen), pD10, psiX174, pBluescript II KS, pNH8A, pNH16a, pNH18A, pNH46A (Stratagene); ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia); Eukaryotic: pSV2CAT, pOG44, pXT1, pSG (Stratagene) pSVK3, pBPV, pMSG, pSVL (Pharmacia). However, any other plasmid or vector may be used as long as they are replicable and viable in the host.

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda P_R, P_L and trp. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

In a further embodiment, the present invention relates to host cells containing the above-described constructs. The host cell can be a higher eukaryotic cell, such as a mammalian cell, or a lower eukaryotic cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-Dextran mediated transfection, or electroporation (Davis, L., Dibner, M., Battey, I., Basic Methods in Molecular Biology, (1986)).

The constructs in host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence. Alternatively, the enzymes of the invention can be synthetically produced by conventional peptide synthesizers.

Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook, et al., Molecular Cloning: A Laboratory

5

10

15

20

Manual. Second Edition, Cold Spring Harbor, N.Y., (1989), the disclosure of which is hereby incorporated by reference.

Transcription of the DNA encoding the enzymes of the present invention by higher eukaryotes is increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp that act on a promoter to increase its transcription. Examples include the SV40 enhancer on the late side of the replication origin bp 100 to 270, a cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers.

5

10

15

20

25

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, e.g., the ampicillin resistance gene of E. coli and S. cerevisiae TRP1 gene, and a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), α-factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated enzyme. Optionally, the heterologous sequence can encode a fusion enzyme including an N-terminal identification peptide imparting desired characteristics, e.g., stabilization or simplified purification of expressed recombinant product.

Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host. Suitable prokaryotic hosts for transformation include <u>E. coli</u>, <u>Bacillus subtilis</u>, <u>Salmonella tvphimurium</u> and various species within the genera Pseudomonas, Streptomyces, and Staphylococcus, although others may also be employed as a matter of choice.

As a representative but nonlimiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from

commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM1 (Promega Biotec, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced by appropriate means (e.g., temperature shift or chemical induction) and cells are cultured for an additional period.

Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents, such methods are well known to those skilled in the art.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, Cell, 23:175 (1981), and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

The enzyme can be recovered and purified from recombinant cell cultures by methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Protein refolding steps can be used, as necessary, in completing

10

15

20

configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps.

The enzymes of the present invention may be a naturally purified product, or a product of chemical synthetic procedures, or produced by recombinant techniques from a prokaryotic or eukaryotic host (for example, by bacterial, yeast, higher plant, insect and mammalian cells in culture). Depending upon the host employed in a recombinant production procedure, the enzymes of the present invention may be glycosylated or may be non-glycosylated. Enzymes of the invention may or may not also include an initial methionine amino acid residue.

 β -galactosidase hydrolyzes lactose to galactose and glucose. Accordingly, the OC1/4V, 9N2-31B/G, AEDII12RA-18B/G and F1-12G enzymes may be employed in the food processing industry for the production of low lactose content milk and for the production of galactose or glucose from lactose contained in whey obtained in a large amount as a by-product in the production of cheese. Generally, it is desired that enzymes used in food processing, such as the aforementioned β -galactosidases, be stable at elevated temperatures to help prevent microbial contamination.

10

15

20

25

These enzymes may also be employed in the pharmaceutical industry. The enzymes are used to treat intolerance to lactose. In this case, a thermostable enzyme is desired, as well. Thermostable β -galactosidases also have uses in diagnostic applications, where they are employed as reporter molecules.

Glucosidases act on soluble cellooligosaccharides from the non-reducing end to give glucose as the sole product. Glucanases (endo- and exo-) act in the depolymerization of cellulose, generating more non-reducing ends (endo-glucanases, for instance, act on internal linkages yielding cellobiose, glucose and cellooligosaccharides as products). β-glucosidases are used in applications where glucose is the desired product. Accordingly, M11TL, F1-12G, GC74-22G, MSB8-6G, OC1/4V, VC1-7G1, 9N2-31B/G and AEDII12RA18B/G may be employed in a wide variety of industrial applications, including in corn wet milling for the separation of starch and gluten, in the fruit industry for clarification and equipment maintenance, in baking for viscosity reduction, in the textile

industry for the processing of blue jeans, and in the detergent industry as an additive. For these and other applications, thermostable enzymes are desirable.

Antibodies generated against the enzymes corresponding to a sequence of the present invention can be obtained by direct injection of the enzymes into an animal or by administering the enzymes to an animal, preferably a nonhuman. The antibody so obtained will then bind the enzymes itself. In this manner, even a sequence encoding only a fragment of the enzymes can be used to generate antibodies binding the whole native enzymes. Such antibodies can then be used to isolate the enzyme from cells expressing that enzyme.

For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler and Milstein, 1975, Nature, 256:495-497), the trioma technique, the human B-cell hybridoma technique (Kozbor et al., 1983, Immunology Today 4:72), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole, et al., 1985, in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96).

Techniques described for the production of single chain antibodies (U.S. Patent 4,946,778) can be adapted to produce single chain antibodies to immunogenic enzyme products of this invention. Also, transgenic mice may be used to express humanized antibodies to immunogenic enzyme products of this invention.

Antibodies generated against the enzyme of the present invention may be used in screening for similar enzymes from other organisms and samples. Such screening techniques are known in the art, for example, one such screening assay is described in "Methods for Measuring Cellulase Activities", *Methods in enzymology*, Vol 160, pp. 87-116, which is hereby incorporated by reference in its entirety.

The present invention will be further described with reference to the following examples; however, it is to be understood that the present invention is not limited to such examples. All parts or amounts, unless otherwise specified, are by weight.

In order to facilitate understanding of the following examples certain frequently occurring methods and/or terms will be described.

10

15

20

"Plasmids" are designated by a lower case p preceded and/or followed by capital letters and/or numbers. The starting plasmids herein are either commercially available, publicly available on an unrestricted basis, or can be constructed from available plasmids in accord with published procedures. In addition, equivalent plasmids to those described are known in the art and will be apparent to the ordinarily skilled artisan.

"Digestion" of DNA refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The various restriction enzymes used herein are commercially available and their reaction conditions, cofactors and other requirements were used as would be known to the ordinarily skilled artisan. For analytical purposes, typically 1 µg of plasmid or DNA fragment is used with about 2 units of enzyme in about 20 µl of buffer solution. For the purpose of isolating DNA fragments for plasmid construction, typically 5 to 50 µg of DNA are digested with 20 to 250 units of enzyme in a larger volume. Appropriate buffers and substrate amounts for particular restriction enzymes are specified by the manufacturer. Incubation times of about 1 hour at 37°C are ordinarily used, but may vary in accordance with the supplier's instructions. After digestion the reaction is electrophoresed directly on a polyacrylamide gel to isolate the desired fragment.

10

15

20

25

Size separation of the cleaved fragments is performed using 8 percent polyacrylamide gel described by Goeddel, D. et al., Nucleic Acids Res., 8:4057 (1980).

"Oligonucleotides" refers to either a single stranded polydeoxynucleotide or two complementary polydeoxynucleotide strands which may be chemically synthesized. Such synthetic oligonucleotides have no 5' phosphate and thus will not ligate to another oligonucleotide without adding a phosphate with an ATP in the presence of a kinase. A synthetic oligonucleotide will ligate to a fragment that has not been dephosphorylated.

"Ligation" refers to the process of forming phosphodiester bonds between two double stranded nucleic acid fragments (Maniatis, T., et al., Id., p. 146). Unless otherwise provided, ligation may be accomplished using known buffers and conditions with 10 units of T4 DNA ligase ("ligase") per 0.5 µg of approximately equimolar amounts of the DNA fragments to be ligated.

Unless otherwise stated, transformation was performed as described in the method of Graham, F. and Van der Eb, A., Virology, 52:456-457 (1973).

Example 1

Bacterial Expression and Purification of Glycosidase Enzymes

DNA encoding the enzymes of the present invention, SEQ ID NOS: 1-14 and 57-60 were initially amplified from a pBluescript vector containing the DNA by the PCR technique using the primers noted herein. The amplified sequences were then inserted into the respective PQE vector listed beneath the primer sequences, and the enzyme was expressed according to the protocols set forth herein. The 5' and 3' primer sequences for the respective genes are as follows:

Thermococcus AEDII12RA -18B/G

5' CCGAGAATTCATTAAAGAGGAGAAATTAACTATGGTGAATGCTATGATTGTC 3' (SEQ ID NO:29)

3' CGGAAGATCTTCATAGCTCCGGAAGCCCATA 5' (SEQ ID NO:30)

Vector: pQE12; and contains the following restriction enzyme sites 5' EcoRI and 3' Blg II.

OC1/4V-33B/G

5' CCGAGAATTCATTAAAGAGGGAGAAATTAACTATGATAAGAAGGTCCGATTTTCC 3' (SEQ ID NO:31)

3' CGGAAGATCTTTAAGATTTTAGAAATTCCTT 5' (SEQ ID NO:32)

Vector: pQE12; and contains the following restriction enzyme sites 5' EcoRI and 3' Bgl II.

Thermococcus 9N2 - 31B/G

5' CCGAGAATTCATTAAAGAGGAGAAATTAACTATGCTACCAGAAGGCTTTCTC 3' (SEQ ID NO:33)

3' CGGAGGTACCTCACCCAAGTCCGAACTTCTC 5' (SEQ ID NO:34)

Vector: pQE30; and contains the following restriction enzyme sites 5' EcoRI and 3' Kpnl.

15

20

Staphylothermus marinus F1 - 12G

5' CCGAGAATTCATTAAAGAGGAGAAATTAACTATGATAAGGTTTCCTGATTAT 3' (SEQ ID NO:35)

3' CGGAAGATCTTTATTCGAGGTTCTTTAATCC 5' (SEQ ID NO:36)

Vector: pQE12; and contains the following restriction enzyme sites 5' EcoRI and 3' Bgl II.

Thermococcus chitonophagus GC74 - 22G

5' CCGAGAATTCATTAAAGAGGAGAAATTAACTATGCTTCCAGGAGAACTTTCTC 3' (SEQ ID NO:37)

3' CGGAGGATCCCTACCCCTCCTCTAAGATCTC 5' (SEQ ID NO:38)

Vector: pQE12; and contains the following restriction enzyme sites 5' EcoRI and 3' BamHI.

MIITL

10

15

20

25

5' AATAATCTAGAGCATGCAATTCCCCAAAGACTTCATGATAG 3' (SEQ ID NO:39)

3' AATAAAAGCTTACTGGATCAGTGTAAGATGCT 5' (SEQ ID NO:40)

Vector: pQE70; and contains the following restriction enzyme sites 5' SphI and 3' Hind III.

Thermotoga maritima MSB8-6G

5' CCGACAATTGATTAAAGAGGAGAAATTAACTATGGAAAGGATCGATGAAATT 3' (SEQ ID NO:41)

3' CGGAGGTACCTCATGGTTTGAATCTCTTCTC 5' (SEQ ID NO:42)

Vector: pQE12; and contains the following restriction enzyme sites 5' EcoRI and 3' KpnI.

Pyrococcus furiosus VC1 - 7G1

5' CCGAGAATTGATTAAAGAGGAGAAATTAACTATGTTCCCTGAAAAGTTCCTT 3' (SEQ ID NO:43)

3' CGGAGGTACCTCATCCCCTCAGCAATTCCTC 5' (SEQ ID NO:44)

Vector: pQE12; and contains the following restriction enzyme sites 5' EcoRI and 3' Kpn I.

Bankia gouldi endoglucanase (37GP1)

- 5' AATAAGGATCCGTTTAGCGACGCTCGC 3' (SEQ ID NO:45)
- 3' AATAAAAGCTTCCGGGTTGTACAGCGGTAATAGGC 5' (SEQ ID NO:46)

Vector: pQE52; and contains the following restriction enzyme sites 5' Bam HI and 3' Hind III.

Thermotoga maritima α-galactosidase (6GC2)

- 5' TTTATTGAATTCATTAAAGAGGAGAAATTAACTATGATCTGTGTGGAAATATTCGGAAAG 3' (SEQ ID NO:47)
- 3' TCTATAAAGCTTTCATTCTCTCACCCTCTTCGTAGAAG 5' (SEQ ID NO:48)

Vector: pQET; and contains the following restriction enzyme sites 5' EcoRI and 3' Hind III.

Thermotoga maritima \(\beta\)-mannanase (6GP2)

- 5' TTTATTCAATTGATTAAAGAGGAGAAATTAACTATGGGGATTGGTGGCGACGAC 3' (SEQ ID NO:49)
- 3' TTTATTAAGCTTATCTTTTCATATTCACATACCTCC 5' (SEQ ID NO:50)

Vector: pQEt; and contains the following restriction enzyme sites 5' Hind III and 3' EcoRI.

AEPII 1a B-mannanase (63GB1)

- 5' TTTATTGAATTCATTAAAGAGGAGAAATTAACTATGCTACCAGAAGAGTTCCTATGGGGC 3' (SEQ ID NO:51)
- 3' TTTATTAAGCTTCTCATCAACGGCTATGGTCTTCATTTC 5' (SEQ ID NO:52)

Vector: pQEt; and contains the following restriction enzyme sites 5' Hind III and 3' EcoRI.

OC1/4V endoglucanase (33GP1)

- 5' AAAAAACAATTGAATTCATTAAAGAGGAGAAATTAACTATGGTAGAAAGACACTTCAGATATGTTCTT
 3' (SEQ ID NO:53)
 - 3' TTTTTCGGATCCAATTCTTCATTTACTCTTTGCCTG 5' (SEQ ID NO:54)

5 .

10

15

20

Vector: pQEt; and contains the following restriction enzyme sites 5' BamHI and 3' EcoRI.

Thermotoga maritima pullalanase (6GP3)
5' TTTTGGAATTCATTAAAGAGGAGAAATTAACTATGGAACTGATCATAGAAGGTTAC 3'
(SEQ ID NO:55)
3' ATAAGAAGCTTTTCACTCTCTGTACAGAACGTACGC 5' (SEQ ID NO:56)
Vector: pQEt; and contains the following restriction enzyme sites 5' EcoRI and 3' Hind III.

The restriction enzyme sites indicated correspond to the restriction enzyme sites on the bacterial expression vector indicated for the respective gene (Qiagen, Inc. Chatsworth, CA). The pQE vector encodes antibiotic resistance (Amp^r), a bacterial origin of replication (ori), an IPTG-regulatable promoter operator (P/O), a ribosome binding site (RBS), a 6-His tag and restriction enzyme sites.

10

15

20

25

The pQE vector was digested with the restriction enzymes indicated. The amplified sequences were ligated into the respective pQE vector and inserted in frame with the sequence encoding for the RBS. The ligation mixture was then used to transform the E. coli strain M15/pREP4 (Qiagen, Inc.) by electroporation. M15/pREP4 contains multiple copies of the plasmid pREP4, which expresses the lacI repressor and also confers kanamycin resistance (Kan^r). Transformants were identified by their ability to grow on LB plates and ampicillin/kanamycin resistant colonies were selected. Plasmid DNA was isolated and confirmed by restriction analysis. Clones containing the desired constructs were grown overnight (O/N) in liquid culture in LB media supplemented with both Amp (100 ug/ml) and Kan (25 ug/ml). The O/N culture was used to inoculate a large culture at a ratio of 1:100 to 1:250. The cells were grown to an optical density 600 (O.D.600) of between 0.4 and IPTG ("Isopropyl-B-D-thiogalacto pyranoside") was then added to a final 0.6. concentration of 1 mM. IPTG induces by inactivating the lacI repressor, clearing the P/O leading to increased gene expression. Cells were grown an extra 3 to 4 hours. Cells were then harvested by centrifugation.

The primer sequences set out above may also be employed to isolate the target gene from the deposited material by hybridization techniques described above.

Example 2

Isolation of A Selected Clone From the Deposited genomic clones

A clone is isolated directly by screening the deposited material using the oligonucleotide primers set forth in Example 1 for the particular gene desired to be isolated. The specific oligonucleotides are synthesized using an Applied Biosystems DNA synthesizer. The oligonucleotides are labeled with 32P--ATP using T4 polynucleotide kinase and purified according to a standard protocol (Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, Cold Spring, NY, 1982). The deposited clones in the pBluescript vectors may be employed to transform bacterial hosts which are then plated on 1.5% agar plates to the density of 20,000-50,000 pfu/150 mm plate. These plates are screened using Nylon membranes according to the standard screening protocol (Stratagene, 1993). Specifically, the Nylon membrane with denatured and fixed DNA is prehybridized in 6 x SSC, 20 mM NaH₂PO₄, 0.4%SDS, 5 x Denhardt's 500 μg/ml denatured, sonicated salmon sperm DNA; and 6 x SSC, 0.1% SDS. After one hour of prehybridization, the membrane is hybridized with hybridization buffer 6xSSC, 20 mM NaH2PO4, 0.4%SDS. 500 ug/ml denatured, sonicated salmon sperm DNA with 1x106 cpm/ml 32P-probe overnight at 42°C. The membrane is washed at 45-50°C with washing buffer 6 x SSC, 0.1% SDS for 20-30 minutes dried and exposed to Kodak X-ray film overnight. Positive clones are isolated and purified by secondary and tertiary screening. The purified clone is sequenced to verify its identity to the primer sequence.

Once the clone is isolated, the two oligonucleotide primers corresponding to the gene of interest are used to amplify the gene from the deposited material. A polymerase chain reaction is carried out in 25 μ l of reaction mixture with 0.5 ug of the DNA of the gene of interest. The reaction mixture is 1.5-5 mM MgCl₂, 0.01% (w/v) gelatin, 20 μ M each of dATP, dCTP, dGTP, dTTP, 25 pmol of each primer and 0.25 Unit of Taq

5

10

15

20

WO 98/24799 PCT/US97/22623

polymerase. Thirty five cycles of PCR (denaturation at 94°C for 1 min; annealing at 55°C for 1 min; elongation at 72°C for 1 min) are performed with the Perkin-Elmer Cetus automated thermal cycler. The amplified product is analyzed by agarose gel electrophoresis and the DNA band with expected molecular weight is excised and purified. The PCR product is verified to be the gene of interest by subcloning and sequencing the DNA product. The ends of the newly purified genes are nucleotide sequenced to identify full length sequences. Complete sequencing of full length genes is then performed by Exonuclease III digestion or primer walking.

Example 3

Screening for Galactosidase Activity

Screening procedures for α -galactosidase protein activity may be assayed for as follows:

10

15

20

Substrate plates were provided by a standard plating procedure. Dilute XL1-Blue MRF E coli host of (Stratagene Cloning Systems, La Jolla, CA) to O.D. $_{600}$ = 1.0 with NZY media. In 15 ml tubes, inoculate 200 μ l diluted host cells with phage. Mix gently and incubate tubes at 37 °C for 15 min. Add approximately 3.5 ml LB top agarose (0.7%) containing 1mM IPTG to each tube and pour onto all NYZ plate surface. Allow to cool and incubate at 37 °C overnight. The assay plates are obtained as substrate p-Nitrophenyl α -galactosidase (Sigma) (200 mg/100 ml) (100 mM NaCl, 100 mM Potassium-Phosphate) 1% (w/v) agarose. The plaques are overlayed with nitrocellulose and incubated at 4 °C for 30 minutes whereupon the nitrocellulose is removed and overlayed onto the substrate plates. The substrate plates are then incubated at 70 °C for 20 minutes.

WO 98/24799 PCT/US97/22623

Example 4

Screening of Clones for Mannanase Activity

A solid phase screening assay was utilized as a primary screening method to test clones for \(\beta \)-mannanase activity.

A culture solution of the Y1090-E. coli host strain (Stratagene Cloning Systems, La Jolla, CA) was diluted to O.D.₆₀₀=1.0 with NZY media. The amplified library from Thermotoga maritima lambda gtl1 library was diluted in SM (phage dilution buffer): 5×10^7 pfu/µl diluted 1:1000 then 1:100 to 5×10^2 pfu/µl. Then 8 µl of phage dilution (5×10^2 pfu/µl) was plated in 200 µl host cells. They were then incubated in 15 ml tubes at 37 °C for 15 minutes.

Approximately 4 ml of molten, LB top agarose (0.7%) at approximately 52 °C was added to each tube and the mixture was poured onto the surface of LB agar plates. The agar plates were then incubated at 37 °C for five hours. The plates were replicated and induced with 10 mM IPTG-soaked Duralon-UVTM nylon membranes (Stratagene Cloning Systems, La Jolla, CA) overnight. The nylon membranes and plates were marked with a needle to keep their orientation and the nylon membranes were then removed and stored at 4 °C.

An Azo-galactomannan overlay was applied to the LB plates containing the lambda plaques. The overlay contains 1% agarose, 50 mM potassium-phosphate buffer pH 7, 0.4% Azocarob-galactomannan. (Megazyme, Australia). The plates were incubated at 72 °C. The Azocarob-galactomannan treated plates were observed after 4 hours then returned to incubation overnight. Putative positives were identified by clearing zones on the Azocarob-galactomannan plates. Two positive clones were observed.

The nylon membranes referred to above, which correspond to the positive clones were retrieved, oriented over the plate and the portions matching the locations of the clearing zones for positive clones were cut out. Phage was eluted from the membrane cut-out portions by soaking the individual portions in 500 µl SM (phage dilution buffer) and 25 µl CHCl₃.

10

15

20

25

Example 5

Screening of Clones for Mannosidase Activity

A solid phase screening assay was utilized as a primary screening method to test clones for \(\beta \)-mannosidase activity.

5

10

15

20

25

A culture solution of the Y1090-*E. coli* host strain (Stratagene Cloning Systems, La Jolla, CA) was diluted to O.D. $_{600}$ =1.0 with NZY media. The amplified library from AEPII 1a lambda gtl1 library was diluted in SM (phage dilution buffer): 5×10^7 pfu/µl diluted 1:1000 then 1:100 to 5×10^2 pfu/µl. Then 8 µl of phage dilution (5×10^2 pfu/µl) was plated in 200 µl host cells. They were then incubated in 15 ml tubes at 37 °C for 15 minutes.

Approximately 4 ml of molten, LB top agarose (0.7%) at approximately 52 °C was added to each tube and the mixture was poured onto the surface of LB agar plates. The agar plates were then incubated at 37 °C for five hours. The plates were replicated and induced with 10 mM IPTG-soaked Duralon-UVTM nylon membranes (Stratagene Cloning Systems, La Jolla, CA) overnight. The nylon membranes and plates were marked with a needle to keep their orientation and the nylon membranes were then removed and stored at 4 °C.

A p-nitrophenyl-\(\beta\)-D-manno-pyranoside overlay was applied to the LB plates containing the lambda plaques. The overlay contains 1% agarose, 50 mM potassium-phosphate buffer pH 7, 0.4% p-nitrophenyl-\(\beta\)-D-manno-pyranoside. (Megazyme, Australia). The plates were incubated at 72 °C. The p-nitrophenyl-\(\beta\)-D-manno-pyranoside treated plates were observed after 4 hours then returned to incubation overnight. Putative positives were identified by clearing zones on the p-nitrophenyl-\(\beta\)-D-manno-pyranoside plates. Two positive clones were observed.

The nylon membranes referred to above, which correspond to the positive clones were retrieved, oriented over the plate and the portions matching the locations of the clearing zones for positive clones were cut out. Phage was eluted from the membrane cut-out portions by soaking the individual portions in 500 μ l SM (phage dilution buffer) and 25 μ l CHCl₃.

5

10

15

20

25

Example 6

Screening for Pullulanase Activity

Screening procedures for pullulanase protein activity may be assayed for as follows:

Substrate plates were provided by a standard plating procedure. Host cells are diluted to O.D. $_{600}$ = 1.0 with NZY or appropriate media. In 15 ml tubes, inoculate 200 μ l diluted host cells with phage. Mix gently and incubate tubes at 37 °C for 15 min. Add approximately 3.5 ml LB top agarose (0.7%) is added to each tube and the mixture is plated, allowed to cool, and incubated at 37 °C for about 28 hours. Overlays of 4.5 mls of the following substrate are poured:

100 ml total volume

0.5g	Red Pullulan Red (Megazyme, Australia)
1.0g	Agarose
5ml	Buffer (Tris-HCL pH 7.2 @ 75 °C)
2ml	5M NaCl
5ml	CaCl ₂ (100mM)
85ml	dH ₂ O

Plates are cooled at room temperature, and thenm incubated at 75°C for 2 hours. Positives are observed as showing substrate degradation.

Example 7

Screening for Endoglucanase Activity

Screening procedures for endoglucanase protein activity may be assayed for as follows:

1. The gene library is plated onto 6 LB/GelRite/0.1% CMC/NZY agar plates (~4,800 plaque forming units/plate) in E.coli host with LB agarose as top agarose. The plates are incubated at 37°C overnight.

- 2. Plates are chilled at 4°C for one hour.
- 3. The plates are overlayed with Duralon membranes (Stratagene) at room temperature for one hour and the membranes are oriented and lifted off the plates and stored at 4°C.
- 4. The top agarose layer is removed and plates are incubated at 37°C for ~3 hours.
 - 5. The plate surface is rinsed with NaCl.
 - 6. The plate is stained with 0.1% Congo Red for 15 minutes.
 - 7. The plate is destained with 1M NaCl.
- 8. The putative positives identified on plate are isolated from the Duralon membrane (positives are identified by clearing zones around clones). The phage is eluted from the membrane by incubating in $500\mu l$ SM + $25\mu l$ CHCl₃ to elute.
- 9. Insert DNA is subcloned into any appropriate cloning vector and subclones are reassayed for CMCase activity using the following protocol:
- i) Spin 1ml overnight miniprep of clone at maximum speed for 3 minutes.
- ii) Decant the supernatant and use it to fill "wells" that have been made in an LB/GelRite/0.1% CMC plate.
 - iii) Incubate at 37°C for 2 hours.
 - iv) Stain with 0.1% Congo Red for 15 minutes.
 - v) Destain with 1M NaCl for 15 minutes.
 - vi) Identify positives by clearing zone around clone.

Numerous modifications and variations of the present invention are possible in light of the above teachings and, therefore, within the scope of the appended claims, the invention may be practiced otherwise than as particularly described.

10

15

20

25

WHAT IS CLAIMED IS:

- 1. An isolated polynucleotide selected from the group consisting of:
 - (a) SEQ ID NOS: 1-14 and 57-60;
 - (b) SEQ ID NOS: 1-14 and 57-60, wherein T can also be U;
 - (c) polynucleotide sequences complementary to SEQ ID NOS: 1-14 and 57-60;
 - (d) polynucleotide sequences which encode an amino acid sequence as set forth in SEQ ID NOS:15-28, and 61-64; and
 - (e) fragments of (a), (b), (c) or (d) that are at least 15 consecutive bases in length and that will selectively hybridize to DNA which encodes a polypeptide of SEQ ID NOS:15-28, and 61-64.
- 2. A vector comprising a polynucleotide of claim 1.
- 3. A host cell containing the vector of claim 2.
- 4. The method of claim 3, wherein the host cell is a eukaryotic cell.
- 5. The method of claim 3, wherein the host cell is a prokaryotic cell.
- 6. A method for producing a polypeptide comprising:
 - (a) culturing the host cells of claim 3;
 - (b) expressing from the host cell of claim 3 a polypeptide encoded by said polynucleotide; and
 - (c) isolating the polypeptide.

PCT/US97/22623

WO 98/24799

- 7. An enzyme selected from the group consisting of:
 - (a) an enzyme comprising an amino acid sequence set forth in SEQ ID NOS: 15-28 or 61-64; and
 - (b) an enzyme which comprises at least 30 consecutive amino acid residue as an enzyme of (a).
- 8. An enzyme of which at least a portion is coded for by a polynucleotide of claim 1, and which is selected from the group consisting of:
 - (a) an enzyme comprising an amino acid sequence which is at least 70% identical to an amino acid sequence selected from the group of amino acid sequences set forth in SEQ ID NOS:15-28 or 61-64; and
 - (b) an enzyme which comprises at least 30 amino acid residues to the enzyme of (a).
- 9. A method for generating glucose from soluble cell oligosaccharides comprising contacting a sample containing oligosaccharides with an effective amount of an enyzme selected from the group consisting of an enzyme having the amino acid sequence set forth in SEQ ID NOS: 15-28, 61-63 and 64 such that glucose is produced.
- 10. The method of cliam 9, wherein the sample is selected from the group consisting of dairy products, fruit juices, detergents, textiles, guar gum, animal feed, plant biomass and waste products.
- The method of claim 9, wherein the oligosaccharide is selected from the group consisting of maltose, cellobiose, lactose, sucrose, raffinose, stachyose, verbascose, cellulose, starch, amylose, glycogen, disacharrides, polysacharrides and pullulan.

M11TL GLYCOSIDASE - 29G COMPLETE GENE SEQUENCE - 9/95

	, , ,	1: AA	.Α <i>:</i>									-			,,,,						
1	H.	1 1.7	2: 114			A GAL	יירין.	. V.Lı	· AT	1 1 1 2 2 A	. T'A1	J. 1. V	ור דר	. 11.V							_
6.1			, ,,,	(22	: ** ***																
.: 1	GI	y 11	e Pri	n Gly		C1	LIAT		. 447	ווא י	CAT	, MX:	TY2;	(ITA	TCK;	CTI	('A'	T (:A	*		:
•											•			V-7-1	Trn	VAI					
12:1			^ L.L./	1377	. ~~ .																
4 1	۸sı	ı Thi	. A }.	Ala	GIV	Leu	Val	Ser	***	1.77	777	CCT.	(:V(VVC	CCC	CCA	(27	TA	. 172		180
														7211	617	PTO	C: 1 v	• • • • • • •			
181		, ,,,,,			~~~	~ ~ ~															
61	Later	128	u (C) u	Asn	Asp	His	Asp	Leu	Ala	Glu	Lve	100	CCC	CIT	AAC	VCT	VII	VC	CT	GCC:	240
241											• -		~ . ,	vai	45 D	Thr	7 l 🗪	A			80
81	• • •	-		AC:T	100																
•••	V4.	610	irp	Ser	YLG	Ile	Phe	Pro	Lys	Pro	Thr	Phe	Asn	Val	Live	UTC	CCT	GTA	CVC	AGA	300
301	GAT	CAC											•••••		ey s	AT	Pro	Val	Clu	YLA	100
101	Asp	Glu	Aen	Clu	AGC	ATT Ile	CIT	CYC	CTA	GAT	CTC	CAT	GAT	AAA	CCC	ست	~~~				
				U1 y	Ser	Ile	Val	His	Val	Asp	Val	Asp	λsp	Lys	Ala	Val	GI.	AGA	CIT	CAT	360
361	222	1117	CCC	AAC	110	~															120
121	Clu	Leu	Ala	λεn	Lve	CAG Glu	GCC (GTA	AAC.	CAT	TAC	CTA	GAA .	ATG '	TAT .	wa '	GAC	TGG	CTT	C) \	
	٠.													uer .	IYE .	Lvs	y ev	T	11-1	~ .	. 420
421 141		CCT	YCY	AAA	-		~~~														140
141	Arg	Gly	Arg	Lys	Leu	Ile i	Leu /	~. \		7	CAT	TGG (500 (CTG (CTC '	TCC	CTT	CAC	AAC	480
										-		·		rea i	LO 1	ren .	Tro	Len	Wie -		160
481	CCA	ATC	ATG	GTG .	ACA .	AC .															
161	PTO	Ile	Het	Val.	Arg /	Arg h	tet (ly i	Pro A	Asp /	Are A	1) a 5	- LL _ 1	CA C	GC 1	rcc (TT.	AAC.	GAG	GAG	. 540
541	~~~		·										10 2	er u	ıy 7	LD I	eu .	λεη	Glu	Glu	180
181	100	GTG	CTC (CAC :	III (GCC A	L AN	AC C	see d	GCA 1	AC A	TT C	כד ז	YCC: 3					_		
	361	0 4 I	VAI	Glu 1	Phe)	Ala L	ys T	yr <i>i</i>	Na A	lla 7	yr I	le A	la T	יו כבי	ve M	125 (7GC (CAC	CTA	CCT	600
601	GTT	ATC	**											-, -	,		TY (310	Leu	Pro	200
201	Val	Met '	700	5	ACC J	LTG A	YC C	XX C	:00 4	MC 0	TC G	T T	AT G	AG C	λλ G	GA T	.ac :				
	Val :			ser ,	Inr E	iec y	sn G	lu P	LO Y	sn V	al V	al T	yr G	lu G	ln G	lv T	// / // /	ior i	Dhe 1	51-1·	660
001	~~ (355 (CCT	בער כ	'C & C							•									220
661 221	Lys (Gly (Gly F	Phe P	TO P	er G	CC T	AC T	TG A	CTT	TG G	AA C	CL C	בד כו	AT A	AG G	CC A	GG ;	CA J	U.T	720
												~	- A	TW VI	ED L	VS A	א או	TO 1			240
721 .	ATG)	NTC (CAG C	כד כ	AT C	C > C															
241	Het 1	ile (Cln A	la H	is A	la A	rg Al	la T	VE A	SD A:	EN T	17 N	u c	SC TO	וכ אנ	א דכ	AG, A	مبر د	CT 0	TT	780
781 (_					,	/3 A2	g r	ie S	er Ly	's L	ys P	TO. V	a l	260
	CCY C	א גד.	TA T	yc c	CL L	לכ כז	UA TO	C T	rc G	AA C:	TA TT	ra ga	C CC	- CC							
261 (SIY L	eu I	lle T	AL Y	la P	pe C)	n T	TO PI	ne G	lu Le	u Le	eu Cl	u G1	. V Pr		A G	W C	TA T	77 C	AT	840
841 2	AG T	T-T 3		~~ ~														al b	ve Y	SÞ	280
281 1	Lys P	he L	.ve c	OC T(CT A	AC TT	'A TA	C TA	T T	רכ אכ	A GA	C AT	'A GT	A TC	G AA	G GC	TAC	T T	C2 2.	~~	000
	ys P		,,	er 36	er L)	/s Le	עד ט	r Ty	'I Ph	e Th	r Aş	p Il	e Va	1 Se	r Ly	5 G1	v Se	er 5	er T	1.	900 300
301 W	TC X	AT G	77 C	እል ተደ	AC AC						_										300
301 I	le A	sn V	al C	lu T	/E Ar	70 AC	7 J.	1 (1	7 60		T AG	C CT	A GA	CTG	C II	c cc	C C:	TA	AC TA	NC	960
							_					2 26	U A5	PIT	P Le	u Gl	v Va		- T	,-	320
301 T	AT A	שב כו	CT 17	CA CT	T 7	C		~ ~			_										
321 T	yr S	er A	rg Le	eu Va	1 Ty	F Ly	s Ile	e Va	l As	D As	n I.v.	n cc.	r AT	A ATO	CI	G CA	c cc	C T	T GC	A :	1020
1021 T											- -,		3 110	= 116	Le:	u Hi	s Gl	y Ty	I C)	Y :	340
341 0	TC CT	T T(ST AC	CC CC	TCC	C CC	3 ATC	: AG	כ ככו	G GC:	T GA	A AA1		-			·				
341 P	ne Le	u cy	ys Th	ır Pr	o C1	A C12	/ Ile	Se:	r Pro	o Álá	G1:	J AST	Pro	Cvs	Set		, 1-1-	TGG	G TC	:C]	1080
1081 G	AG GT	Y: T.	* ~											-,-			<i>y</i>	e 61	утг	P :	60
361 G	lu Va	1 1	,, LC	4 GV	^ CC.	A CTC	TAC	יודם :	CT	נדט ז	w	CAA	CTI	TAC		. CG1	TA4	כיהר	ت ت		140
	lu Va	- ,		J ()	u GI	y Leu	Tyr	Lei	ı Lei	, Leu	: Lys	Clu	Leu	Tyr	Asr	Arc	TV	נים.	u Us	~ i	140
ridi Cy	*C 7T	C AT	יני כדי	C Arr																	
381 As	AC Tr	u II	e Va) Tie	- 5/1.		CLO	U-1	TCA	GAC	ACC	ACT:	GAT	CCC	TTC	AGA	ccc	cc.	A TA	C 1	200
							•				., -, -,	AI G	V21)	V19	Leu	Arg	Pro	3 A1	a Tv	. 4	00
1201 64	TO GT	(: TC	C CA.	T COM																	
401 Le:	u Va	F Se	r His	s vai	Tyi	Ser	Val	T'r n	Lue	Als	OC T	YVC.	CIVC	CCC.	ATT	CCC	CTC	·	N GG(: 1	260
									-,		~	A::01	(1143	CHY	He	Pro	Va I	1.0		, 4	20
TAUL TA	ני ניתו		1. 323	: 4/1																	
AN LA	r læs	. Hi:	: T',	See	1.000	The	Asp	A2:11	Tyr	Glu	Time	Λ I	C.L.	CALC.	1.14.	VC:C:	CAG	: ^^	יורר ו	. 1	1.20
											•			y	1-130-	Arg	Glin	1.y:	i Ph	4 -	40

Figure la

1 (#)		172;	· :Ac:	A74.									. ,,	1.4.61	۸٠٠,	r		۸۱.,	1.041) 4344 444)
463 1441 481	L.VC.	AAT	14	146	Ala	The	His	Asia	132A 131Y	110	Pro	СИТ Дир	GAG Glu	(TFA Li)	r'Act Gla	r'AT Hass	i TPP Lani	AC'A T711	i Tri: Israi	VJA.	1440 480

Figure 1b(Continued)

WO 98/24799

OC1/4 GLYCOSIDASE - 33G/B COMPLETE GENE SEQUENCE - 9/95

ATT: ATA AGA ACC TOTAL
Het Tie Arg Arg For TTT (T'A AAA CAT TTT ATC TTC CC)
1 ATG ATA AGA AGG TOT GAT TIT GTA AAA GAT TIT ATC TIT GGA ACG GGT AGG GGA TAC 60 61 GAG ATT GAA GGT GGA AGG GAA AAG GAA GGB Phe Pin Lys Aup Phe Ile Phe Gly Thr Alm Thr Alm Alm Tyr 20
" I NG ATT CAR com an
Gin Tie Glu Gly Ala Ala ANG GAT GGC AGA GGG CCA TCA ATT TOTAL
21 Gin Tie Giu Gly Ale Ale Ash Gilu Asp Gily Arg Gly Pro Ser Tie Try Asp Val Phe Ser 40
TAT CAC ACC CCT con
41 His The Pro Gly Lys The Leu Ash Gly Asp The Gly Asp Val Ala Cys Asp His Tyr His 60
181 CGA TAG AND THE GIY ASP THE GIY ASP VAL ALB CYS ASP HIS TYP HE
THE CON TAC AND CAR OF THE EN
ATG TYP LYS Glu ASP Ile Glo Lou ATG AAA GAA ATA GGG TTA GAC GCT TAG AGG
61 Arg Tyr Lys Glu Asp Ile Gin Leu Net Lys Glu Ile Gly Leu Asp Ala Tyr Arg Phe Ser 80
81 116 FOR TOC TOC CCC AGA ATT ATC CCA GAT GGG ANG AND
Ser TIP PIO ATG Ile Het PIO ATG CAN AAC ATC AAC CAN AAC GGT GTG CAN AAC
101 The arm of the Asn Cln Lys Clv Val Asn TC 300
101 TAC ANC AGA CTC GTT GAT GAG CTT TTG ANG AND
301 TAC AAC AGA CTC GTT GAT GAG CTT TTG AAG AAT GAT ATC ATA CCA TTC GTA ACA CTC TAT 360 101 Tyr Asn Arg Leu Val Asp Glu Leu Leu Lys Asn Asp Ile Ile Pro Phe Val Thr Leu Tyr 120
361 CAG TOO Phe Val Thr Lev 198 Asn Asp Ile Ile Pro Phe Val Thr Lev 198
361 CAC TGG GAC TTA CCC TAC GCA CTT TAT GAA AAA GGT GGA TGG CTT AAC CCA GAT ATA GCG 420
121 His TIP ASP Leu Pro Tyr Ala Leu Tyr Glu Lys Gly Gly Trp Leu Asp Pro Asp Ile Ala 140
421 CTC me Asn Pro Asn Ile 110
141 LEU TAT TTC AGA GCA TAC GCA ACG TTT ATG TTC ALG
421 CTC TAT TTC AGA GCA TAC GCA ACG TTT ATG TTC AAC GAA CTC GGT GAT CGT GTG AAA CAT 141 Leu Tyr Phe Arg Ala Tyr Ala Thr Phe Het Phe Asn Glu Leu Gly Asp Arg Val Lys His 481 TGG ATT ACA CTG AAC GAA GCA TGG TGG AAC CAT GTG AAC CAT GBG TGG ATG TGG AAC CAT GBG TGG AAC CAT GBG TGG AAC CAT GBG TGG TGG AAC CAT GBG TGG TGG TGG TGG TGG TGG TGG TGG TG
481 TOO ATT LOS ATT LOS
481 TGG ATT ACA CTG AAC GAA CCA TGG TGT TCT TCT TCC GGT TAT TAC ACG GGA GAG CAT 540
161 TEP Ile The Leu Ash Glu Pro Tep Cys Ser Ser Phe Ser Gly Tyr Tyr The Gly Glu His 180
541 GCC CCC ccm are a second s
181 Ala Pro Civi Mar TTA CAA GAA GCG ATA ATC GCC COD DIA
541 CCC CCG GGT CAT CAA AAT TTA CAA GAA GCG ATA ATC GCG GCG CAC AAC CTC TTG AGG GAA .600 181 Ala Pro Gly His Gln Asn Leu Gln Glu Ala Ile Ile Ala Ala His Asn Leu Leu Arg Glu 200 601 CAT GGA CAT GCC GTC CAG GCG TCC AGA GCG TCC AGA GAA ATC GCG GCG CAC AAC CTC TTG AGG GAA .600
501 CAT CCA can are
201 His Gly His Ala Val Gln Ala Ser Arg Glu Glu Val Lys Asp Gly Glu Val Gly Leu Thr 220
Ala Val Gln Ala Ser Arg Glu Glu Val Lya Arn CCG GAX GTT GGC TTA ACC 660
661 AAC CTT CMC 200
661 AAC GTT GTG ATG AAA ATA GAA CCG GGC GAT GCA AAA CCC GAA AGT TTC TTG GTC GCA AGT 720 721 CTT GTT GAT AAG TTC GTT AAR GGA CCG GGC GAT GCA AAA CCC GAA AGT TTC TTG GTC GCA AGT 720 721 CTT GTT GAT AAG TTC GTT AAR GGA CCG GGC GAT GCA AAA CCC GAA AGT TTC TTG GTC GCA AGT 720
The GIU Pro Gly Asp Ala Lys Pro Glu Ser Ple TTG GTC GCA AGT 720
721 CTT GTT GAT ANG TTC GTT AND GGO 240
721 CTT GTT GAT AAG TTC GTT AAT GCA TGG TCC CAT GAC CCT GTT GTT TTC GGA AAA TAT CCC 780 781 GAA GAA GCA GTT GCA CTT TAN AGA
ASH Ala Trp Ser His Asp Pro Val Val Pho Cly AND TAT CCC 780
781 GAA GAA GCA GTT GCA CTT TAT ACC COLUMN
781 GAA GAA GCA GTT GCA CTT TAT ACC GAA AAA GGG TTG CAA GTT CTC GAT AGC GAT ATG AAT 840 841 ATT ATT TCG ACT CCT ATA CAG TTG TTG TTG TTG TTG TTG TTG TTG TTG T
17- THE GIU LYS Gly Leu Gln Val Leu Asp Ser Ash Mer AT 840
ATT ATT TCG ACT CCT ATA GAC TTC TTT CCT ATA
841 ATT ATT TCG ACT CCT ATA GAC TTC TTT GGT GTG AAT TAT TAC ACA ACA ACA CTT GTT GTT 900 901 TTT GAT ATG AAC AAT CCT GTT GTT GTT GTT GTT GTT GAT ATG AAC ATG ATG AAC AAT CCT GTT GTT GTT GTT GAT ATG AAC AAT CCT GTT GTT GTT GTT GTT GAT ATG AAC AAT CCT GTT GTT GTT GTT GTT GTT GAT ATG AAC AAT CCT GTT GTT GTT GTT GTT GTT GAT ATG AAC AAT CCT GTT GTT GTT GTT GTT GTT GTT GTT GT
901 TIT GAT ATC AAC AND
901 TIT GAT ATG AAC AAT CCT CTT GGA TIT TCG TAT GTT CAG GGA GAC CTT CCC AAA ACG GAG 960
AME HET ASH ASH Pro Leu Gly Phe Ser Tyr Val Gly Gly Che Cor AMA ACG GAG 960
961 ATG GCA money and 120
961 ATG GGA TCG GAA ATC TAC CCG CAG CGA TTA TTT GAT ATG CTG GTC TAT CTG AAG GAA AGA 1020
The Giu He Tyr Pro Gin Gly Leu Phe Asp Het Leu Val The AAG GAA AGA 1020
1021 TAT AAA CON SOLUTION SOLU
341 Tyr Lys Leu Pro Leu Tyr Ile Thr Glu Ash Gly Het Ala Gly Pro Asp Lys Leu Glu Ash 360
Det Fib Leu Tyr Ile Thr Glu Asn Gly Het Ala Gly Pro Ash AAA TIG GAA AAC 1080
1081 GGA AGA COM
1081 GGA AGA GTT CAT GAT AAT TAC CGA ATT GAA TAT TTC GAA AAG CAC TTT GAA AAA GCA CTT 1140 1141 GAA CCA ATC AAT GCA GAT GTT CAT GAT GAT GAT GAT GAT GAT GAT GAT GAT G
ASP ASP TYP Arg Ile Glu Tyr Leu Glu Lys His Phe Clu Lys His Ph
1141 GAA CCA amp 380
381 Glu Ala Ile ASD Ala ASD VILLANT TTG AAA GCT TAC TTG ATT TGG TGT TTG
ASP Val ASP Leu Lys Cly Tyr Phe Ile TID Ser Leu Mes ACC 1200
1201 TTC GAA TEC CON TOO 400
Phe Glu Trp Ala Cys Gly The Cot ATA ATC TAC CTA CAT THE CAT TH
1761 On the last type of the Cly IIe IIe Tyr Val ASD Tor ANT ACC 1260
1261 CCA AAA AGG ATA TRE AAA GAT TCA GCG ATG TGG TTG AAG GAA TTT CTA AAA TCT TAA 1317 421 Pro Lys Arg He Len Lys Asp Ser Ala Het Trp Len Lys Glu Pho Lys Asp Tct TAA 1317
421 Pro Lys Arg He Len Lys Asp Ser Ala Met Trp Leu Lys Glu Phe Leu Lys Ser End 419
The Lev Lys Glu Phe Lev Lys Ser End 419
·

Figure 2

STAPHYLOTHERMUS MARINUS GLYCOSIDASE COMPLETE GENE SEQUENCE 9/95

1 TTG ATA AGG TIT CCT GAT TAT TTC TTG TIT GGA AGA GGT AGA TCA TGG GAG CAG ATY:	
1 MEE I LE AND PHE PRO ASP TYP PHE LEU PHE GLY THE ALA THE SET SET HIS GLO LIFE GLY THE ALA TAK AND AND THE SET SET HIS GLO LIFE GLY THE ALA TAK AND AND THE SET SET HIS GLO LIFE GLY THE ALA TAK AND AND THE SET SET HIS GLO LIFE GLY THE ALA TAK AND AND THE SET SET HIS GLO LIFE GLY THE ALA TAK AND THE SET SET HIS GLO LIFE GLY THE ALA TAK AND THE SET SET HIS GLO LIFE GLY THE ALA TAK AND THE SET SET HIS GLY THE ALA TAK AND THE SET SET SET HIS GLY THE ALA TAK AND THE SET SET SET SET THE ALA TAK AND THE SET SET SET SET THE SET SET SET SET SET SET SET SET SET SE	
Tyr Phe Leu Phe Gly Thr Ala Thr Ser Ser His Cle ATT	11A1: 60
61 GGT AAT AAC ATA TIT AAT GAT TGG TCC CAG TCC	610 2 0
61 CGT AAT AAC ATA TIT AAT GAT TGG TGG GAG TGG GAG AGT AAA GGC AGG ATT AAG GTG A	ica una
21 Gly Asn Asn Ile Phe Asn Asp Trp Trp Glu Trp Glu Thr Lys Gly Arg Ile Lys Val /	VCA. 120
121 TCG CCT AAG GCA TCT AAT CAT TGG GAA CTC TAT AAA GAA GAC ATA GAG CTT ATG GCT C 41 Ser Gly Lys Ala Cys Asn His Trp Glu Leu Tyr Lys Glu Asp Ile Gly Lou	, 40
41 Ser Gly Lys Ala Cys Asn His Trp Glu Leu Tyr Lys Glu Asp Ile Glu Leu Het Ala G	AG 180
181 CTG GGA TAT AND GG	lu so
61 Leu Gly Tyr Ash Ale The Tot TCC ATA GAG TGG AGT AGA ATA TOTAL	
61 Leu Gly Tyr Asn Ala Tyr Arg Phe Ser Ile Glu Trp Ser Arg Ile Phe Pro Arg Lys A. 241 CAT ATA GAT TAT GAG TOC COM NO.	እፕ 240
241 CAT ATA GAT TAT GAG TCG CTT AAT AAG TAT AAG GAA ATA GTT AAT CTA CTT AGA AAA TA Bl His Ile Asp Tyr Glu Ser Leu Asn Lys Tyr Lys Glu Ile Val Asn Lou Land Ada Aa	SP 80
BI His Ile ASP Tyr Glu Ser Leu ASP TAG GAA ATA GTT AAT CTA CTT AGE AND	
81 His Ile ASP TYR GAG TCG CTT AAT AAG TAT AAG GAA ATA GTT AAT CTA CTT AGA AAA TA 301 GGG ATA GAA CCT GTA ATC ACC	AC 300
301 GGG ATA GAA CCT GTA ATC ACT CTT CAC CAC TTC ACA AAC CCG CAA TGG TTT ATG AAA AT 101 Gly 1le Glu Pro Val 1le Thr Leu His His Phe Thr Asn Pro Gly Top The Tax AAA AT	T 100
101 Gly Ile Glu Pro Val Ile Thr Leu His His Phe Thr Asn Pro Gln Trp Phe Het Lys Il	T 360
361 GGT GGA TCC Acc acc	e 120
J61 GGT GGA TGG ACT AGG GAA GAG AAC ATA AAA TAT TTT ATA AAA TAT GTA GAA CTT ATA GC 121 Gly Gly Trp Thr Arg Glu Glu Asn Ile Lys Tyr Phe Ile Lys Tyr Val Glu Leu Ile Al 121 TCC GAG ATA AAA GAG GTG All Ash GAG Ash GAG All Ash GAG Ash	
And the Giu Giu Asn Ile Lys Tyr Phe Ile Lys Tyr Val Cit ATA AC	T 420
421 TCC GAG ATA AAA GAC GTG AAA ATA TGG ATC ACT ATT AAT GAA CCA ATA ATA TAT GTT TT: 141 Ser Glu Ile Lys Asp Val Lys Ile Trp Ile Thr Ile Asp Glu Pro Ile Tat GTT TT:	a 140
141 Ser Glu Ile Lys Asp Val Lys Ile Trp Ile Thr Ile Asn Glu Pro Ile Ile Tyr Val Let 481 CAA GGA TAT ATT TGG GGG GAA GAA ATA TAT GTT TT	
481 Cla Pro Ile Tyr Val Le	A 480 160
481 CAA GGA TAT ATT TCC GGC GAA TGG CCA CCT GGA ATT AAA AAT TTA AAA ATA GGT GAT CAA	- 100
161 Gln Gly Tyr Ile Ser Gly Glu Trp Pro Pro Gly Ile Lys Asn Leu Lys Ile Aia Asp Gln GTA ACT ANG ANT CTT TTO AND AND GLO GAT CAN	540
541 GTA ACT ALC AND GENERAL ASP GIR	180
541 GTA ACT ANG ANT CIT TTA ANA GCA CAT ANT GNA GCC TAT ANT ATA CIT CAT ANA CAC GGT Val Thr Lys Asn Leu Leu Lys Ala His Asn Glu Ala Tyr Asn Ile Leu His Lys His Gly ATT GTA GGC ATA GCT ANA CAC GGT	
Ded Ley Ala His Asn Glu Ala Tyr Asn Ile Leu His Inc.	600
601 ATT GTA GGC ATA GCT ANA AAC ATG ATA GCA TTT ANA CCA GGA TCT AAT AGA GGA ANA GAC 201 Ile Val Gly Ile Ala Lys Asn Het Ile Ala Phe Lys Pro Gly Ser her AGA GGA ANA GAC	200
201 Ile Val Gly Ile Ala Lys Asn Het Ile Ala Phe Lys Pro Gly Ser Asn Arg Gly Lys Asp 661 ATT AAT ATT TAT CAT Ala Company	660
661 ATT 12m are Cly Lys Asp	220
661 ATT AAT ATT TAT CAT ANA GTC GAT ANA GCA TTC AAC TGG GGA TTT CTC AAC GGA ATA TTA 221 Ile Asn Ile Tyr His Lys Val Asp Lys Ala Phe Asn Trp Gly Pha Lys	
221 Ile Asn Ile Tyr His Lys Val Asp Lys Ala Phe Asn Trp Gly Phe Leu Asn Gly Ile Leu 721 AGG GGA GAA CTA GAA AGG GGA TGG GGA TTA GAA GGA ATA TTA	720
721 ACG GGA GAA CTA GAA ACT CTC CGT GGA AAA TAC CGA GTT GAG CCC GGA AAT ATT GAT TTC	240 .
241 Arg Gly Glu Leu Glu Thr Leu Arg Gly Lys Tyr Arg Val Glu Pro Gly Asn Ile Asp Phe 781 ATA GGC ATA AAC TAT THE COLUMN TO THE ASP Phe	700
The state of the s	780 260
781 ATA GGC ATA AAC TAT TAT TCA TCA TAT ATT GTA AAA TAT ACT TGG AAT CCT TTT AAA CTA 261 Ile Gly Ile Asn Tyr Tyr Ser Ser Tyr Ile Val Lys Tyr Thr TTD Acc CCT TTT AAA CTA	200
THE GIY HE ASH TYP TYP Ser Ser TYP HE VALUE AND THE TOT GAST CCT TIT AND CTA	840
841 CAT ATT AND COO PAGE 150 Per Lys Leu	280
841 CAT ATT ANA GTC GAA CCA TTA GAT ACA GGT CTA TGG ACA ACT ATG GGT TAC TGC ATA TAT 281 His Ile Lys Val Glu Pro Leu Asp Thr Gly Leu Trp Thr Thr Het Gly Tyr Cys Ile Tyr 901 CCT AGA GGA ATA TAT CAL CTAT	
of the Fro Leu Asp Thr Gly Leu Trp Thr Thr Het Gly Tre Con Th	900
901 CCT AGA GGA ATA TAT GAA GTT GTA ATG AAA ACT CAT GAG AAA TAC GGC AAA GAA ATA ATC 301 Pro Arg Gly Ile Tyr Glu Val Val Het Lys Thr His Glu Lys Tyr Gly Lys ATA ATC	300
301 Pro Arg Gly 1le Tyr Glu Val Val Het Lys Thr His Glu Lys Tyr Gly Lys Glu Ile Ile	960
961 ATT ACC COLOR OF THE SIZE	320
961 ATT ACA GAG AAC GGT GTT GCA GTA GAA AAT GAT GAA TTA AGG ATT TTA TCC ATT ATC AGG	•
121 Ile Thr Glu Asn Gly Val Ala Val Glu Asn Asp Glu Leu Arg Ile Leu Ser Ile Ile Arg	1020
1021 CAC TTA CAR THE SET ITE ILE Arg	340
141 His Leu Gin Tyr Leu Tyr Live All GCA ATG AAT GAA GGA GCA AAG GTG AAA GGA TAT TTC TIC	
141 His Leu Gln Tyr Leu Tyr Lys Ala Het Asn Glu Gly Ala Lys Val Lys Gly Tyr Phe Tyr	1080 360
166 AGC TTC ATC GAT AAT TIT GAG TGG GAT AAA CO.	200
1081 TGG AGC TTC ATC GAT AAT TTT CAG TGG GAT AAA GGA TTT AAC CAA AGG TTC GGA CTA GTA 161 Trp Ser Phe Het Asp Asn Phe Glu Trp Asp Lys Gly Phe Asn Gln Arg Phe Gly Leu Val	1-140
1141 GAA CTT CAM DID AND AND AND AND AND AND AND AND AND AN	380
381 GIU VAI ASD TYF LY THE GAG AGA AAA CCT AGA AAA AGC CCA TAT CTA	
THE PHE CLU ARE LYS PRO ARE LYS SET ALA THE LYS CAN	1200
1201 ATA UCA COT ACC AAG ACT ATA ACT GAT GAA TAC CTA GAA AAA TAT GGA TTA AAG AAC CTC	400
401 THE ALB ARG THE LYS THE THE SEE ASP GIU TYE LEU GIU LYS TYE GIY LEU LYS ASE LEU 1261 GAA TAA 1766	1260
1761 Try Leu Lys Asn Leu	1200
421 (1) -	
Ciu End 422	-

Figure 3

Thermococcis 9N2 Glydesidase -318/0 Complete gene sequence 9/95

		AT	C						•		•		10-110	.e y	75						
		l Ne	E La	2 Pc4	· cl.	GGC	LIL	CIC	TCC	್ಯ	ברה:	TCC	CAG	700	CC		·	_		ATG (
						CIA	Phe	Leu	מזב	Cly	Val	Ser	Gln	Sar	E1 v	Pho		TTC	GAG	ATG (CC 60
	6	LCA	دنمه		ACC	**			_						,		OLA	rne	Glu	SAT C	37A 30
	2:	λø	P Ly	Leu	Aru	A 5.5	AAC	A	GAT	. כנים	AAC	ACA	CAC	TCG	TCC	AAC.	705			CAT C	
•			•			~~ 0	Ain	110	Vab	1.0	AJD	Thr	Asp	Trp	Tip	Lvz	Tm	מוכ	AGG	CYI C	120
	121	12	CAAC	ATA	AAG	ACC	GA N	~~								-, -	,	• • • •	AIG	ASD F	Pro 40
	41	Ph	a Amr	ile	Lye	AED	Clv	CTC	CTC	VCC.	CCC	CYC	(CCC	CAC	CAC	CCC	4-4		AAC I	
					-				A7 I	Ser	cly	γap	Leu	2ro	Clu	Clu	Clv	11.	~~	AAC T	75 18D
	181	ردی	, c11	TÁC	CAC	AAG	CAT	ChC	F/~.	~-·.									~4D	۲ مهد مهد	YF 60
	61	Cli	Leu	TYI	Glu	Lys	ARD	W1 =	1	Cic	CCC	بمنتد	CYC	CTC	CCI	CTG	λλC	CIT	TAC	AGG A	
						•			A. 6	Leu	7-4	YLâ	yab	Leu	Gly	Leu .	Asso.	Val	TVI	Ara T	77 240
	341	CCY	ATA	CYC	766	AGC .	YCC .	ATC											- 3 - .	-y 1.	le 80
	81	Cly	11=	CJA	Top	Ser .	Arg .	Ile	Ph.	PTO	7	TCA	YCC .	;GG	111	CIC :	CAG (CIT :	صد ا	val el Arg I	
	201								- 1.7-	,,,	.rp	FEO	722	ئ تن	3pe .	wal (Glu '	Val.	Aero 1	/al ci	300
	201	CGG	CYC	YCC	TAC	GGA 1	cre d	STC .	WC.	CAC	G PP			•		~				/al 61	In 100
	TOT	YLâ	yeb	Ser	34.	Cly :	Lou 1	/al	Lys	Ase	Tal :	nca tua	ATC (GAT .	AAA (באכ ז	vce (322	:AA (iAG CI	760
	161		-:-			٠.						Uy s	114 /	(CE	LYS)	do :	pr I	let (ilu (lu	14 120
	17:	m/C	CAC	ATA	500	WT (CAT (:XC (EAC .	ATA	ככר י	T1.C	T							TC AC	- 110
		ASD	CTF	ile .	, פונא	(22)	tia C	in c	21.1	Tie .	Aia 1	·~	~~ (-60	-GC (777)	EX C	inc :	عدد	דכ אם •ע אב	C 420
	421	GAG	~~									• , •	, 7L ,	7 B 1	rab /	/41 1	10 0	lu i	lis L	של טי	9 140
	141	Glo	7	CL	(סבד	uc c	TC X	יבי כ	TC 2	MC (uc (CAC T	-TC 3	~ ~					**	
			~40	CIA	720 I	ys V	'al I	1. v	741)	Nan 1	د ری	un !	dia P	5 u 7	46 C	C	בכ כ	TC T	, 200 C	TT CA	C 480
4	481	GAT	ccc	171	-										71E L		TO L	au T	LD T	eu Hi	160
2	161	λED	P=0	71A 7	Cic c	CC Y	.cc c	yc y	ye c	:C: (TC A	CC A	uc c	CT A	CC 1	7T A				eu Hi: ≈ CX	
			0	-14 1	TO Y	TE Y	tā C	lu L	ye i	ila :	.eu T	nr A	sn C	lvi	ro T	14 0	SC T	CC C	IC C	jλ C]; ≈ cγ	\$ 540
3	341	CAC	محد ا	ene e	TC -	\^ =											1 y 17	ע פב	er c	JA CJ	180
1	.81	Glu	Ser 1	Val v	'A1 C	AL T	TC C	∞ Y	ag t	JYC C	<u> </u>	CG 1	אכ אי	TC G	CG A	.	~	~ ~		ib Fer nc clid là Cli	•
2	01	Val .	rab i	let T	TD 50	er m	hr Di		AC 5	AG C	== Y	<u> </u>	-c c		rc &	× 0	T 00	77 T	. ~	n Per C CC: N Per	
2:	21	Pro :	TYT S	er G	ly Pi	le Pr	O PY			rr a	אסד	יב מ	≖ 0	4	35 CC	עגב	ري عا	~ GC	A 17	e Fen C C1C	
- 4	. 1	Yau ;	et I	- A	וא ביו	a Hi	3 Al	A Le	וגניי	-C 11	~ ~	u, A7	TX CT	א אי	- N	3 11	COX	C AC	8 63.	عدد	780
	•	VIO Y	SP L	YE XS	₽ 5e	ד אב	g Se	r Gl	נג ט	a G1	u Va	- c	N AZ	A AI	C 11	= XX	c w	CAT	A CC	Val	840
																					280
	-		y = 7:	e ry	T 7.5	P . Se:	T Asi	ر مرد د	p Fr	o Ly	3 45	D Va	1 12	- A1:	- ~		1. 33 60	. כא	CYYC	לאן דער	900
90	1 7	TC C	20 20										,		- ~	- G11	1 75	י ארי) Acc	12.	300
30	1 1	he P	ie ce	,	G C.T		1110	GA.	: 60	A AT	C CA	CAN	s ccc	. AAC		- 33/				CAC	
																					960
9 5																					320
32	1 0	Jy 6:	lu Th	T Pha	ובט	7	<u> </u>		CA) ACC	CC) AAC	عدد	TCC	ATE	. con	مبنی ،		TAC	
																					1020
																					340
241	L	YI TH	I Ar	g Glu	1 Val	VAI	A	7.A.T	100	GAC	. 000	AAC	TTC	co	AGC	ATA	. 000	CIG	172	TCC	1000
																					1080 360
1081																					300
361	. 4	e or ne yr	9 C1	y Val	His	Am	TVE	63.4	C	7.	יים בי	ACC	ccc	œc	ACT	101	TCC	GCC	CAC	GGA	1140
											-	•		,	344	307	-5 - 2	A 1 -	\	C 3 · · ·	380
1141		~ ~ ~				2-0			_												
787	λı	G Pr	o Vaj	Scr	Asp	Ile	Cly	TIP	61.7	11-	Tur	Per	C	CCC	ATC	TAC	CYC	TCG	ATA	AGA	1200
1201																					400
401	_		- 22				_														•
	4 1	u Ala	A Auto	Lys	12.2	Cly	Va i	Pro	Va 1	Tyr	Val	The	63	A.C.	C	ATA	ccc	CXT	TCA	λCT	1260
1261										-	_			~~~	VIY	110	414				420
421			10	CTAC	~~	~			_												
-		ולד כ	~= 11	vīđ	PIO	IAL	JÄL	Lou	Ala	Ser	His	Val	Ala	LYB	11	Cla	~~	CCC	TAC	CYC	1320
													-		~~ ~	21 U	-10	WT.	TYE	Glu	440

Figure 4a

1321 441 .381	Ala	61	T TA(GAC	GT	ACC	Coc	- TAC	: CT:	TAC	. TGG	GCC.								GCC	
	•	•	.,.	~=;	, va.	. YEC	Gly	יגלי. י	Lon	· 1/1	Tro	Ala	1	700	GAC	YYC	TAC	CAG	100	CCC	1380
. 381	c_{CC}	CCT					•								~-,	VE1	1,3,2	ີເປັນ	مين	414	410
461	Leu	Gly	Pho		ATC	AGG	220	. eec	CTC	TAT	222	CTC:	CAT						·		7.0
461				~29	net	yrg	Phe	CIA	Len	Tyr	Lys	Val	ART	Lau	ATA	YCC	WC	CAC	AGA	Ala	1440
1441	CCC	· ccc	CAC				•									4115	LY 5	G) u	Ara	fhr	410
481	Pro	YES	Glu	Clu	ALC:	CTA	ALC.	CIL	747	AGC	CCC	ATC	CTC	C10							
		-		414	361	-VAI	Lys	Yal	Tyr	λrg	GLy	I .u	Val	01	VVC	WC	CCY	o c	ACC	MC	1500
150;	CAN	ATC	CCC			_								OT.	VID.	Ven.	CIA	Va)	Ser	Lys	500
501	Clu	11.	Ara	Glu	~10	-10	CCA	CĻI	CRG	TGA	15	30								•	
				010	-y=	Pha	Cly	Len	Glv	End	91	Α.									

Figure 4b(Continued)

	1		GAA				AT (GAA A	TT (כדכ	זרד	(,VC	17.	A A	СT	ACA	GAG	: CA	۸ ۸۸	G (7)					
	,	Mei			r Ik					L2-6	Scr	Cin	l.cv	T	hr	Thr	Cin	Glu	n nn Lys				TC ev	V#1	60 20
	61 21	CTC Val	GCG		TO			CA (GA (111	CCC	۸۸	c c	CA	CAT	TCC	λG	· care	; uc	C C	ா ம	cc	CCT	120
	21		GAA		•	-						Cly	Asn			His	Ser	Arg	Val	۸la	Gi		h	Ala	40
		Gly			His		n Y	AT C	CA A	NGA Nrg	CTT Leu	GGA Gly	AT:	T C		GCG Na	TTT Phe	CTC Vai	: כדכ			T G	CT .	ccc	130
1	81	GCA	GGA	CT	C AC	A AT	ΓΑ Α			_	AGG	•							Leu TAC	دا ۸	A.s.	p G	ly	Pro	60
	61	A la	Gly	Leu	Ar	e lle	٨	sn P	o T	July 1	Arg	Glu	Asq	. G/	P C	JAA Jiu	AAC Asn	ACT	TAC	TAI Tyr	_		CG	GCA	240
24	41 ^{- 1} 31 - 1	П	ccc	CTT	r GA	A A7	C Y.	TG C	rc c	ст :	rct	ACC	TGG		\C A	GA	GAC			-				Ala	80
						•	171		~		cr	Thr	Τmp	عه	n A	TZ.	Asp	Lev	Lev	GA, Glu	A GA Glu		_	GGA Gly	300 100
30)	.уг	GCC Ala	ATO	G GG	A GA	بر ان دا	W 6	<u> </u>	GG (344	TAC	GGT	CI	c a	AT	CTG	CIT	СТТ	GCA	· cc	· T GC		ATG	360
36				•	•			- '	• •		,,,,	ı yı	Giy	Val		₽ .	Val	Lev	Lev	Ala	Pro			Met	120
12	1 /	usn I	le	His	. AG	A AA Asr	C CC	ים ה שלים	T TO	ज ०	iGA .	AGG Arg	AAT Am	TT(C G				TCA	GAA	GA:	r cc	т (яc	420
42	, ,	11 1	rcc	CCT	GA.	A AT	G GC		•		-	-			_		Tyr	Tyr	Ser CAA	Glu	Asp	Pto	, '	/ai	140
14	1 · L	=u 5	er	Gly	Glu	Me	Ali	Ser	Al	P	he '	√a)	Lys	Gly	۸ G ۷	17 (CAA Gla	TCT Ser	CAA Gin	GGG	GT C	G GG		GCC da	480
48 161		GC A				ונו		C GC	ن م	.c .	AE (AG	GAA	AC	5	AC A	.cc	ATC		GTG		-			160
	_			Lyx		Phe	-		~-		C	ann .	CIN	i hr	٨s	a 🖊	VZ .	Ma	Val	Val	GA C	AC:		TC e	540 180
54 I 181	ν.	IG T	et CC	GAG Glu	CGA Arg	GC(CT.	C AG	A GA	کر کر 10 ادا	T A T	AT	<u> </u>	<i></i>	G	ו ה		GAA	ATT	GCT	صو	**	5 A		600
601	G(5A A										ут						Glu	lle	Ala	Val	Lys		ys	200
201	Al	2 A	ř <u>z</u>	Pro	Trp	ACC Thr	Val	Met	Ser	C GO	न T • T	AC /	AAC Asn	AAA Lys	CT Les	G A		GGA Gly	AAA Lys	TAC	ाटा	TCA		AG	660
661 221		\C G	М :	TGG	СТТ	TTG Leu	**	3 446	; cr	ו כו	. Y								-	. Tyr	Cys	Ser	G	n	220
				-			-,-	-,,	* #1	LE	. ~	rg C	مزاد	Clu	Τπρ	· C	ly I	Pac	GGC	COT	TTC Phe	GTG VeJ	A' M	rg _{Et}	720 240
721 241	A C Ser	C G,	4C 7	TOG	TAC	GC5 Ala	GGA	GAC Asp	AA	c cc	т с	TA C	A A	CAG	СТ	ء ک	AG (cc	GGA	MC	GAT	ATG	A7		780
781		G CC		•	•		٠.,	~up	V21	rto	٠ ٧	11 C	ilu	Gin	Leu	Ļ	/S /	\la	Gly	Asn	Азр	Mei	ile		260
261		Pro		ily I	Lys	GCG Ala	Tyr	CAG Gln	Val	AA Asn	C AC	CA G	AA lu	AGA Arg	AG/	A GA				GAA	GAA	ATC	AT	-	840
841	GA	G GC	G T	TG /	WG	GAG Glu											•		ite	Clu	Glu	He	M	1	280
231					•		٠.,	Cys	LEU	261	Gli	i G	י עו	Vai	Leu	. GA As		AG (TGT Cys	CTG Val	AGA Arz	AAC Asn	AT lic	-	900 300
30 I	رين احا	Lys	A (7	TT C	TT .	GTG Val	AAC Asn						44 (CCG	TAC	AG	G T	AC 7	rca .	AAC	AAG	ccc	GA	•	960
961		-							Pro	Ser	Pho	: Ly	's (Sly	Tyr	۸rį			er	Asn	Lys	Pro	Asp		900 020
321	Leu	Glu	Sc	r H	is is	GCG Ala	GAA Giv	GTC Val	GCC Ala	TAC	GA	A G(.A (GT Hy			G G			crc	CTT	CTT	GA	5 11	020
1021	AAC	AÁC	: GC	n c	т	CTT	cca			•				-	Ala	Clo		y V	'al	Vai	Lev	Lru	Clu		940
341	Asn	Asn	Ci	y V:	j)				4.4	CIU	∧z¤	Thi	11	4	Val	Alo	v.	. 10		GGĆ GIY	ACC Thr	CCT	CAA)80)60
391 1081	ATC	GAA	AC	A A	ΓΑ ,	AAG	GCA	GGA .	\CG	GGA	٨٥٦	GG	A G	۸C.	ACC	CAT				•		Gly	Gin		
							-	•	•••	٠.,	Jei	City	^.	ų,	The	His	. Pru	, ,	rg -		ACG Thr	ATC He	TCT Ser		40 80
1141 381		Lev	GA.	A GC GF	, i ;t. '	ATA . le l	A A A	GAA /	GA	AAC	ATG	**	c n	רני ו	GVC.	GAA	\ GA	۸ ر	דכ מ	·	TCC	۸ίΤ	FAT	12	DX)
						. '	,.	Clu A	· · · ·	VZU	Mei	l.yv	Ph	ie ,	Амі	Glu	Ch					The	Tyr		00

Figure:5a

1201 GAG GAG TAC ATA AAA AAG ATG AGA GAA ACA GAG GAA TAT AAA CCC, AGA VC.C CAC F(7 TGG Glu Glu Tyr IIc Lys Lys Met Arg Glu Thr Glu Glu Tyr Lys 401 1260 Pro AIE l'br Τm 4.20 1261' GGA ACG GTC ATA AAA CCG AAA CTC CCA GAG AAT TTC TCA CAG AAG 421 Gly Thr 444 Val lic Lys 132n Pres Lys Lev Pro Cin Scr Glu Lyx Lys 440 1321 CCT CCA AAG AAA AAC GAT כוד פנא כוד כוד כום ATC AGT AGG ATC TCC CCT GAG GGA 441 TAC Pro Pro Lys Lys Asp Val Ala 1380 Val Val lic -Ser Arg lic Gly Clu Cly Tyr 460 1381 GAC AGA AAG CCG GTG AAA GGT GAC TTC TAC CTC TCC GAT GAC GAG CTG GAA Asp Arg Lys Pro Val CTC ATA Lys 1440 Gly Asp Pho Tyr Leu Ser A.sp Glu وعا He Lys 480 1441 ACC GTC TCG AAA GAA TTC CAC GAT CAG GGT AAG AAA GTT GTG CTT CIT CTG Thr Val Ser Lys Glu Phe AAC ATC GGA His Asp Gin Gly Lys 1500 Lys Val Val Leu He Gly 500 1301 AGT CCC ATC GAA GTC GCA AGC TGG AGA GAC CTT GTG GAT GGA ATT CTT CTC CTC Ser Pro Ile Glu Val Ala Ser Trp Arg Asp Leu TGG CAG 1560 Val Gly He . Leu Val Trp Gin 520 1561 GCG GGA CAG GAG ATG GGA AGA ATA GTG GCC GAT GTT CTT GTG GGA ANG Ala Gly ATI CCC Gin Glu Met TCC Gly 1620 Arg Ile Val Αb Αsp Val Val Gly Asn Pro Scr 540 1621 GGA AAA CTT CCA ACG ACC TTC CCG AAG GAT TAC TCG GAC GIT CCA TCC TGG 541 Gly Lys Leu Pro Thr ACG TTC CCA Thr 1680 Phe Pro Lys Αф Tyr Ser Αsp Val Рто \$ 27 Тηр Thr Pnc Pro 560 1681 GGA GAG CCA AAG GAC AAT CCG CAA AGA GTG GTG TAC GAG GAA GAC ATC 561, Gly Gle Pro Lys TAC CTG GGA TAC 1740 Azn Pro Gin Arg Val Val Tyr Glu Clu Αsp Тут City , Tyr 580 1741 AGG TAC TAC GAC ACC TTC GGT GTG GAA CCT GCC TAC GAA TTC 581 Arg Tyr GGC TAC GGC CTC TCT TAC Tyr Asp Thr 1800 Phe Gly Val Giu Pro Тут Clu Pho Gly Tyr Gly Ser Tyr 600 ACA ANG TIT GAA TAC AMA GAT TTA AMA ATC GCT ATC GAC GGT GAG ACG CTC Thr Lys AGA Phe Glu Tyr CTG TCG 1860 Lys Asp Leu Lys lic lte Thr Asp Gly Glu Lev Arz Val Ser 620 TAC ACG ATC ACA AAC ACT GGG GAC AGA GCT GGA AAG GAA GTC 621 Tyr Thr lie Thr Asn TCA CAG OTO TAC ATC Thr 1920 Gly Asp Arg Ala Gly Lys Glu Val Ser Val Gin Tyr He Lys 640 1921 GCT CCA AMA GGA AMA ATA GAC AMA CCC TTC CAG GAG CTG AAA Pro Lys Gly Lys GCG TI CAC ACA 1980 ** lle ** Asp Lys Pro Phe Gìn Glυ Lev Lvs His Lys Thr Lys 660 CTT TTG AAC CCG GGT GAA TCA GAA GAA ATC TCC TTG GAA ATT 661 CCT CTC AGA GAT GCG CTT 2040 Asn Pro Gly Giu Ser Glu Clu He Ser Leu Cin He Pro يرعط Arg ASD Lev Ala 680 AGT TTC GAT GGG AAA GAA TGG GTT GTC GAG TCA GGA GAA TAC 2041 GAG 180 AGG GTC CCT GCA 2100 Asp Gly Lys Glu Trp Vat Val Clu Ser Gly Glu Tyr Giv Arz Val City Ala 700 2101 TCT TCG AGG GAT ATA AGG TTG AGA GAT ATT TTT CTG GTT GAG GGA GAG AAG AGA Arg Asp TTC 2160 Arg Leu Arg Asp He Val Giu Gly Glu Lys 720 Arg Lys 2161 CCA TGA 2166 721 Pro End 722

Figure 56(Continued)

THERMOCOCCUS AEDII12RA GLYCOSIDASE ATC ATC CAC TOC CCG GTT AAA CCG ATT ATA TCT GAG GCT CGC GGC ATA AUC ATC ACA ATA Het lie His Cys Pro Val Lys Gly He lie Ser Glu Ala Arg Gly He Thr He Thr He 61 CAT TTA AGT TTT CAA GGC CAA ATA AAT TTG GTG AAT GCT ATG ATT GTC TTT ECG GAG 20 Asp Leu Ser Phe Gln Gly Gln Ile Asn Asn Leu Val Asn Ala Met Ile Val Phe Pro Glu 120 121 TTC TTC CTC TTT GGA ACC GCC ACA TCT TCT CAT CAG ATC GAG GGA GAT AAT AAA TGG AAC 40 41 Phe Phe Leu Phe Gly Thr Ala Thr Ser Ser His Gln Ile Glu Gly Asp Asn Lys Trp Asn 180 CAC TOG TOG TAT TAT GAG GAG ATA COT AAG CTC CCC TAC AAA TCC GGT AAA GCC TGC AAT 60 Asp Trp Trp Tyr Glu Glu Ile Gly Lys Leu Pro Tyr Lys Ser Gly Lys Ala Cys Asn 240 CAC TOG GAG CTT TAC AGG GAA GAT ATA GAG CTA ATG GCA CAG CTC GGC TAC AAT GCC TAC RO. His Trp Glu Leu Tyr Arg Glu Asp Ile Glu Leu Het Ala Gln Leu Gly Tyr Asn Ala Tyr 300 301 CGC TIT TCG ATA GAG TGG AGC CGT CTC TTC CCG GAA GAG GGC AAA TTC AAT GAA GAA GCC 100 Arg Phe Ser Ile Glu Trp Ser Arg Leu Phe Pro Glu Glu Gly Lys Phe Asn Glu Glu Ala 360 TTC AAC CGC TAC CGT GAA ATA ATT GAA ATC CTC CTT GAG AAG GGG ATT ACT CCA AAC GTT 120 Phe Asn Arg Tyr Arg Glu Ile Ile Glu Ile Leu Leu Glu Lys Gly Ile Thr Pro Asn Val 420 421 ACA CTG CAC CAC TTC ACA TCA CCG CTG TGG TTC ATG CGG AAG GGA GGC TTT TTG AAG GAA 141 Thr Leu His His Phe Thr Ser Pro Leu Trp Phe Het Arg Lys Gly Gly Phe Leu Lys Glu 480 GAA AAC CTC AAG TAC TGG GAG CAG TAC GTT GAT AAA GCC GCG GAG CTC CTC AAG GGA GTC 160 Clu Asn Leu Lys Tyr Trp Glu Gin Tyr Val Asp Lys Ala Ala Glu Leu Leu Lys Gly Val . 540 541 ANG CTT GTA GCT ACA TTC AND GAG CCG ATG GTC TAT GTT ATG ATG GGC TAC CTC ACA GCC 180 181 Lys Leu Val Ala Thr Phe Asn Glu Pro Het Val Tyr Val Het Het Gly Tyr Leu Thr Ala 600 501 TAC TOG CCC CCC TTC ATC AAG AGT CCC TTT AAA GCC TTT AAA GTT GCC GCA AAC CTC CTT Tyr Trp Pro Pro Phe Ile Lys Ser Pro Phe Lys Ala Phe Lys Val Ala Ala Asn Leu Leu 201 660 ANG GCC CAT GCA ATG GCA TAT GAT ATC CTC CAT GGT AAC TIT GAT GTG GGG ATA GTT ALA 663 220 Lys Ala His Ala Het Ala Tyr Asp Ile Leu His Gly Asn Phe Asp Val Gly Ile Val Lys 221 770 AND ATO COO ATA ATO CTO COT GOA AGO AND AGA GAG ANA GAG GTA GAA GOT GOO CAA ANG 240 241 Asn Ile Pro Ile Het Leu Pro Ala Ser Asn Arg Glu Lys Asp Val Glu Ala Ala Gln Lys 780 GCG CAT AAC CTC TIT AAC TGG AAC TTC CTT GAT GCA ATA TGG AGC GGA AAA TAT AAA GGA Ala Asp Asn Leu Phe Asn Trp Asn Phe Leu Asp Ala Ile Trp Ser Gly Lys Tyr Lys Gly B40 GCT TIT GGA ACT TAC ANA ACT COA GAA AGC CAT GCA GAC TIC ATA GGG ATA AAC TAC TAC 841 280 Ala Phe Gly Thr Tyr Lys Thr Pro Glu Ser Asp Ala Asp Phe Ile Gly Ile Asm Tyr Tyr 900 ACA GCC AGC GAG GTA AGG CAT AGC TGG AAT CCG CTA AAG TIT TTC TTC GAT GCC AAG CTT 300 Thr Ala Ser Glu Val Arg His Ser Trp Asn Pro Leu Lys Phe Phe Phe Asp Ala Lys Leu 961 GCA GAC TTA AGC GAG AGA AAA ACA GAT ATG GGT TGG AGT GTC TAT CCA AAG GGC ATA TAC 320 Ala Asp Leu Ser Glu Arg Lys Thr Asp Het Gly Trp Ser Val Tyr Pro Lys Gly Ile Tyr 321. 1020 1021 GAA GCT ATA GCA AAG GTT TCA CAC TAC GGA AAG CCA ATG TAC ATC ACG GAA AAC GGG ATA 340 Glu Ala Ile Ala Lys Val Ser His Tyr Gly Lys Pro Het Tyr Ile Thr Glu Asn Gly Ile 1080 GCT ACC TTA GAC GAT GAG TGG AGG ATA GAG TIT ATC ATC CAG CAC CTC CAG TAC GIT CAC 360 361 Ala Thr Leu Asp Amp Glu Trp Arg Ile Glu Phe Ile Ile Gln His Leu Gln Tyr Val His 1140 ANA GCC TTA ANC GAT GGC TIT GAC TTG AGA GGC TAC TTG TAT TGG TCT TIT ATG GAT ANC 1141 380 Lys Ala Leu Asn Asp Gly Phe Asp Leu Arg Gly Tyr Phe Tyr Trp Ser Phe Net Asp Asn 1200 1201 THE GAG TEG CET GAG CET TIT ACA CEA CEC TIT GEG CTG GTC GAG TAC ACG ACE 400 Phe Glu Trp Ala Glu Gly Phe Arg Pro Arg Phe Gly Leu Val Glu Val Asp Tyr Thr Thr TTC ANG AGG AGA CCG AGA ANG AGT GCT TAC ATA TAT GGA GAA ATT GCA AGG GAA ANG ANA 1261 420 Phe Lys Arg Arg Pro Arg Lys Ser Ala Tyr Ile Tyr Gly Glu Ile Ala Arg Glu Lys Lys 1320 1321 ATA ANA GAC GAA CTG CTG GCA ANG TAT GGG CTT CCG GAG CTA TGA 440 441 Ile Lys Asp Glu Leu Leu Ala Lys Tyr Gly Leu Pro Glu Leu End

Figure 6

THERMOCOCCUS CHITONOPHAGUS GLYCOSIDASE - 22G COMPLETE SEQUENCE - 9/95

1 TTC CTT CCA CAG AAC TTT CTC TGG GGA GTT TCA CAG TCC GGA TTC CAG TTT GAA ATC GUG 1 Het Leu Pro Glu Asn Phe Leu Trp Gly Val Ser Gln Ser Gly Phe Gln Phe Glu Het Gly 2	
The new IPD GIV VAL Car Classes	_
Ser Gly Phe Gly Mer Cly	50
61 GAC AGA CTG AGG AGG CAC ATT GAT CCA AAC ACA GAT TGG TGG TAC TGG GTA AGA GAT GAA 1 21 ASP Arg Leu Arg Arg His Ile ASP Pro Asn Thr ASP Trp Trp Typ Typ Typ Typ Typ Typ Typ Typ Typ Ty	30
21 ASP Arg Leu Arg Arg His Ile ASP Pro Asn Thr Asp Trp Trp Tyr Trp Val Arg Asp Glu 4	
THE ASP PTO ASR THE ASP TTP TYP TYP Val Arg ASP COM	20
121 TAT AAT ATC AAA AAA GGA CTA GTA AGT GGG GAT CTT CCC GAA GAC GGT ATA AAT TCA TAT 1 41 Tyr Asn Ile Lys Lys Gly Leu Val Ser Gly Asp Leu Pro Gly Asp Cly 22	0
41 Tyr Asn Ile Lys Lys Gly Leu Val Son Col GAT CTT CCC GAA GAC GCT ATA AAT TCA TAT	
41 Tyr Asn Ile Lys Lys Gly Leu Val Ser Gly Asp Leu Pro Glu Asp Gly Ile Asn Ser Tyr 61	80
181 GAA TTA TAT GAG AGA GAC CAA GAA ATT GCA AAG GAT TTA GGG CTC AAC ACA TAT AGG ATC 26	U
61 Glu Leu Tyr Glu Arg ASD Gln Glu Tla All ANG GAT TTA GGG CTC AAC ACA TAT AGG ATC	
61 Glu Leu Tyr Glu Arg Asp Gln Glu Ile Ala Lys Asp Leu Gly Leu Asn Thr Tyr Arg Ile 80	40
241 GGA ATT GAA TGG AGC AGA GTA TTT CCA TGG CGA AGG	
241 GGA ATT GAA TGG AGC AGA GTA TIT CCA TGG CCA ACG ACT TIT GTC GAC GTG GAG TAT GAA 30 81 Gly Ile Glu Trp Ser Arg Val Phe Pro Trp Pro Thr Thr Phe Val Asp Val Glu Tyr Glu 10	٠.
301 ATT GAT CAC men and	
301 ATT GAT GAG TCT TAC GGG TTG GTA AAG GAT GTG AAG ATT TCT AAA GAC GCA TTA GAA AAA 36 101 Ile Asp Glu Ser Tyr Gly Leu Val Lys Asp Val Lys Ile Ser Lys Asp Val AAA 36	, 0
THE ASP Glu Ser Tyr Gly Leu Val Lys ASP Val Lys Ile Ser Lyn Ash CAC GCA TTA GAA AAA 36	n
	•
LEU ASP Glu Ile Ala Asp Glu Arg Glu Ile Ile TVT TOT ACC ARC CTA AAA TCC CTA 42	0
141 Arg Lys Arg Gly Phe Lys Val Ile Leu Asn Leu Asn His Phe Thr Leu Pro Ile Trp Leu 160)
481 CAT CAT COT and COT Let 160)
481 CAT GAT CCT ATC GAA TCT AGA GAA AAA GCC CTG ACC AAT AAG AGA AAC GGA TGG GTA AGC 540	
161 His Asp Pro Ile Glu Ser Arg Glu Lys Ala Leu Thr Asn Lys Arg Asn Gly Trp Val Ser 180)
541 GAA AGG AGT COT 121 CO. 120 CO. 180)
541 GAA AGG AGT GTT ATA GAG TTT GCA ANA TTT GCC GCG TAT TTA GCA TAT ANA TTC GGA GAC 600	
TOT ING DEC CO.	
TO ALL DIA GAT ITC TOO NAME AND A PROPERTY OF THE PROPERTY OF	
201 Ile Val ASD Het TIT Ser THE ART GAA CCT ATG GTG GTC GCC GAG TTG GTG TAT TTD	
The val Ala Glu Leu Clu Com to	
THE TAC TREE TREE CON THE CONTRACTOR OF THE CONT	
221 Ala Pro Tyr Ser Gly Phe Pro Pro Gly Val Het Asn Pro Glu Ala Ala Lys Leu Val Het 240	
The sale and the s	
THE CAN LATE ATTE ATT AND COO OFF THE	
241 Leu His Met Ile Asn Ala His Ala Leu Ala Tyr Arg Met Ile Lys Lys Phe Asp Arg Lys 260	
The same and the bys Lvs pha her has the	
AND GET GAT COL CALLERS	
261 Lys Ala Asp Pro Glu Ser Lys Glu Pro Ala Glu Ile Gly Ile Ile Tyr Asn Asn Ile Gly 280	
841 CTC ACL THE CONTROL OF THE CONTR	
841 GTC ACA TAT CCG TTT AAT CCG AAA GAC TCA AAG GAT CTA CAA GCA TCC GAT AAT GCC AAT 900 281 Val Thr Tyr Pro Phe Asn Pro Lys Asp Ser Lys Asp Leu Gle Ala GCA TCC GAT AAT GCC AAT 900	
901 TTC TTC CAC ACT COO TO	
901 TTC TTC CAC AGT GGG CTA TTC TTA ACG GCT ATC CAC AGG GGA ANA TTA AAT ATC GAA TTT 960	
THE WAY LIVE LIPE AND THE CALL AND	
961 GAC GGA GAC ACA TOTA COM TOTAL OF THE STATE OF THE ST	
961 GAC GGA GAC ACA TOTA COM TOTAL OF THE STATE OF THE ST	
961 GAC GGA GAG ACA TIT GIT TAC CIT CCA TAT TIA AAG GGC AAT GAT TGG CTG GGA GTG AAT 1020 121 ASP Gly Glu Thr Phe Val Tyr Leu Pro Tyr Leu Lys Gly ASP ASP TER Leu Gly Val	
961 GAC GGA GAG ACA TIT GIT TAC CIT CCA TAT TIA AAG GGC AAT GAT TGG CTG GGA GTG AAT 1020 1021 TAT TAT ACA ACA GAR GAR GAR GAR GAR GAR GAR GAR GAR GA	
961 GAC GGA GAG ACA TIT GIT TAC CIT CCA TAT TIA AAG GGC AAT GAT TGG CTG GGA GTG AAT 1020 1021 TAT TAT ACA ACA GAR GAR GAR GAR GAR GAR GAR GAR GAR GA	
961 GAC GGA GAG ACA TIT GIT TAC CIT CCA TAT TIA AAG GGC AAT GAT TGG CTG GGA GTG AAT 1020 121 ASD Gly Glu Thr Phe Val Tyr Leu Pro Tyr Leu Lys Gly Asn AsD Trp Leu Gly Val Asn 1020 1021 TAT TAT ACA AGA GAA GTC GIT AAA TAC CAA GAT CCC ATG TIT CCA AGT ATC CCT CTC ATA 1080 141 Tyr Tyr Thr Arg Glu Val Val Lys Tyr Gln Asp Pro Met Phe Pro Ser Llo Pro Lev Gly Ada 1080	
961 GAC GGA GAG ACA TIT GIT TAC CIT CCA TAT ITA AAG GGC AAT GAT TGG CTG GGA GTG AAT 1020 121 ASP Gly Glu Thr Phe Val Tyr Leu Pro Tyr Leu Lys Gly Asn Asp Trp Leu Gly Val Asn 140 1021 TAT TAT ACA AGA GAA GTC GTT AAA TAC CAA GAT CCC ATG TIT CCA AGT ATC CCT CTC ATA 1080 1081 AGC TTC AAG GGC GTT GGA GAG GAG GTC GTT AGA GAT CCC ATG TTC CCA AGT ATC CCT CTC ATA 1080	
961 GAC GGA GAG ACA TIT GIT TAC CIT CCA TAT ITA AAG GGC AAT GAT TGG CTG GGA GTG AAT 1020 121 ASP Gly Glu Thr Phe Val Tyr Leu Pro Tyr Leu Lys Gly Asn Asp Trp Leu Gly Val Asn 140 1021 TAT TAT ACA AGA GAA GTC GTT AAA TAC CAA GAT CCC ATG TIT CCA AGT ATC CCT CTC ATA 1080 1081 AGC TTC AAG GGC GTT GGA GAG GAG GTC GTT AGA GAT CCC ATG TTC CCA AGT ATC CCT CTC ATA 1080	
961 GAC GGA GAG ACA TIT GIT TAC CIT CCA TAT TIA AAG GGC AAT GAT TGG CTG GGA GTG AAT 1020 121 ASP Gly Glu Thr Phe Val Tyr Leu Pro Tyr Leu Lys Gly Asn Asp Trp Leu Gly Val Asn 140 1021 TAT TAT ACA AGA GAA GTC GTT AAA TAC CAA GAT CCC ATG TIT CCA AGT ATC CCT CTC ATA 1080 1081 AGC TTC AAG GGC GTT CCA GAT TAT GGA TAC GGA TGT AGA CCA GGA ACG ACG TCA AAG GAC 1140 1081 AGC TTC AAG GGC GTT CCA GAT TAT GGA TAC GGA TGT AGA CCA GGA ACG ACG TCA AAG GAC 1140 1081 Ser Phe Lys Gly Val Pro Asp Tyr Gly Tyr Gly Cys Arg Pro Gly Thr Thr Ser Lys Acc	
961 GAC GGA GAG ACA TIT GIT TAC CIT CCA TAT TIA AAG GGC AAT GAT TOG CTG GGA GTG AAT 1020 121 ASP Gly Glu Thr Phe Val Tyr Leu Pro Tyr Leu Lys Gly Asn Asp Trp Leu Gly Val Asn 340 1021 TAT TAT ACA AGA GAA GTC GTT AAA TAC CAA GAT CCC ATG TIT CCA AGT ATC CCT CTC ATA 1080 141 Tyr Tyr Thr Arg Glu Val Val Lys Tyr Gln Asp Pro Het Phe Pro Ser Ile Pro Leu Ile 360 1081 AGC TTC AAG GGC GTT CCA GAT TAT GGA TAC GGA TCT AGA CCA GGA ACG ACG TCA AAG GAC 1140 1081 Ser Phe Lys Gly Val Pro Asp Tyr Gly Tyr Gly Cys Arg Pro Gly Thr Thr Ser Lys Asp 380	
961 GAC GGA GAG ACA TIT GIT TAC CIT CCA TAT TIA AAG GGC AAT GAT TOG CTG GGA GTG AAT 1020 121 ASP Gly Glu Thr Phe Val Tyr Leu Pro Tyr Leu Lys Gly Asn Asp Trp Leu Gly Val Asn 340 1021 TAT TAT ACA AGA GAA GTC GTT AAA TAC CAA GAT CCC ATG TIT CCA AGT ATC CCT CTC ATA 1080 141 Tyr Tyr Thr Arg Glu Val Val Lys Tyr Gln Asp Pro Het Phe Pro Ser Ile Pro Leu Ile 360 1081 AGC TTC AAG GGC GTT CCA GAT TAT GGA TAC GGA TCT AGA CCA GGA ACG ACG TCA AAG GAC 1140 1081 Ser Phe Lys Gly Val Pro Asp Tyr Gly Tyr Gly Cys Arg Pro Gly Thr Thr Ser Lys Asp 380	
961 GAC GGA GAG ACA TIT GIT TAC CIT CCA TAT TIA AAG GGC AAT GAT TGG CTG GGA GTG AAT 1020 121 ASP Gly Glu Thr Phe Val Tyr Leu Pro Tyr Leu Lys Gly Asn Asp Trp Leu Gly Val Asn 340 1021 TAT TAT ACA AGA GAA GTC GIT AAA TAC CAA GAT CCC ATG TIT CCA AGT ATC CCT CTC ATA 1080 1341 Tyr Tyr Thr Arg Glu Val Val Lys Tyr Gln Asp Pro Het Phe Pro Ser Ile Pro Leu Ile 360 1081 AGC TTC AAG GGC GTT CCA GAT TAT GGA TAC GGA TGT AGA CCA GGA ACG ACG TCA AAG GAC 1140 1081 Ser Phe Lys Gly Val Pro Asp Tyr Gly Tyr Gly Cys Arg Pro Gly Thr Thr Ser Lys Asp 380 1141 GGT AAT CCT GTT ACT GAC ATT GGA TGG GAG GTA TAT CCC AAA GGC ATG TAC GAC TCT ATA 1200 181 Gly Asn Pro Val Ser Asp Ile Gly Trp Glu Val Tyr Pro Lys Gly Het Tyr Asp Car The	
961 GAC GGA GAG ACA TIT GIT TAC CIT CCA TAT TIA AAG GGC AAT GAT TGG CTG GGA GTG AAT 1020 121 ASP Gly Glu Thr Phe Val Tyr Leu Pro Tyr Leu Lys Gly Asn Asp Trp Leu Gly Val Asn 140 1021 TAT TAT ACA AGA GAA GTC GIT AAA TAC CAA GAT CCC ATG TIT CCA AGT ATC CCT CTC ATA 1080 141 Tyr Tyr Thr Arg Glu Val Val Lys Tyr Gln Asp Pro Het Phe Pro Ser Ile Pro Leu Ile 360 1081 AGC TTC AAG GGC GTT CCA GAT TAT GGA TAC GGA TGT AGA CCA GGA ACG ACG TCA AAG GAC 1140 161 Ser Phe Lys Gly Val Pro Asp Tyr Gly Tyr Gly Cys Arg Pro Gly Thr Thr Ser Lys Asp 380 1141 GGT AAT CCT GTT AGT GAC ATT GGA TGG GAG GTA TAT CCC AAA GGC ATG TAC GAC TCT ATA 1200 1201 GTA GCT GCC AAT CLA TIT GIA TGG GAG GTA TAT CCC AAA GGC ATG TAC GAC TCT ATA 1200	
961 GAC GGA GAG ACA TIT GIT TAC CIT CCA TAT TIA AAG GGC AAT GAT TGG CTG GGA GTG AAT 1020 121 ASP Gly Glu Thr Phe Val Tyr Leu Pro Tyr Leu Lys Gly Asn Asp Trp Leu Gly Val Asn 140 1021 TAT TAT ACA AGA GAA GTC GIT AAA TAC CAA GAT CCC ATG TIT CCA AGT ATC CCT CTC ATA 1080 141 Tyr Tyr Thr Arg Glu Val Val Lys Tyr Gln Asp Pro Het Phe Pro Ser Ile Pro Leu Ile 360 1081 AGC TTC AAG GGC GTT CCA GAT TAT GGA TAC GGA TGT AGA CCA GGA ACG ACG TCA AAG GAC 1140 161 Ser Phe Lys Gly Val Pro Asp Tyr Gly Tyr Gly Cys Arg Pro Gly Thr Thr Ser Lys Asp 380 1141 GGT AAT CCT GTT AGT GAC ATT GGA TGG GAG GTA TAT CCC AAA GGC ATG TAC GAC TCT ATA 1200 1201 GTA GCT GCC AAT CLA TIT GIA TGG GAG GTA TAT CCC AAA GGC ATG TAC GAC TCT ATA 1200	
961 GAC GGA GAG ACA TIT GIT TAC CIT CCA TAT TIA AAG GGC AAT GAT TGG CTG GGA GTG AAT 1020 321 ASP Gly Glu Thr Phe Val Tyr Leu Pro Tyr Leu Lys Gly Ash Asp Trp Leu Gly Val Ash 340 1021 TAT TAT ACA AGA GAA GTC GTT AAA TAC CAA GAT CCC ATG TIT CCA AGT ATC CCT CTC ATA 1080 141 Tyr Tyr Thr Arg Glu Val Val Lys Tyr Gln Asp Pro Het Phe Pro Ser Ile Pro Leu Ile 360 1081 AGC TTC AAG GGC GTT CCA GAT TAT GGA TAC GGA TGT AGA CCA GGA ACG ACG TCA AAG GAC 1140 1081 Ser Phe Lys Gly Val Pro Asp Tyr Gly Tyr Gly Cys Arg Pro Gly Thr Thr Ser Lys Asp 380 1141 GGT AAT CCT GTT ACT GAC ATT GGA TGG GAG CTA TAT CCC AAA GGC ATG TAC GAC TCT ATA 1200 1381 Gly Ash Pro Val Ser Asp Ile Gly Trp Glu Val Tyr Pro Lys Gly Het Tyr Asp Ser Ile 400 1201 GTA GCT GCC AAT GAA TAT GGA GTT CCT GTA TAC GTA ACA GAA AAC GGA ATA GGA GAT TCA 1260 1201 Val Ala Ala Ash Glu Tyr Gly Val Pro Val Tyr Val Thr Glu Ash Gly Ile Ala Ala Ala Ash Glu Tyr Gly Val Pro Val Tyr Val Thr Glu Ash Gly Ile Ala Ala Ala Ala Ash Glu Tyr Gly Val Pro Val Tyr Val Thr Glu Ash Gly Ile Ala Ala Ala Ala Ash Glu Tyr Gly Val Pro Val Tyr Val Thr Glu Ash Gly Ile Ala Ala Ala Ala Ash Glu Tyr Gly Val Pro Val Tyr Val Thr Glu Ash Gly Ile Ala Ala Ala Ash Glu Tyr Gly Val Pro Val Tyr Val Thr Glu Ash Gly Ile Ala Ala Ala Ash Gly Ile Ala Ash Gly Il	
961 GAC GGA GAG ACA TIT GIT TAC CIT CCA TAT TIA AAG GGC AAT GAT TGG CTG GGA GTG AAT 1020 1221 ASD Gly Glu Thr Phe Val Tyr Leu Pro Tyr Leu Lys Gly Asn Asd Trd Leu Gly Val Asn 1020 1021 TAT TAT ACA AGA GAA GTC GIT AAA TAC CAA GAT CCC ATG TIT CCA AGT ATC CCT CTC ATA 1080 1341 Tyr Tyr Thr Arg Glu Val Val Lys Tyr Gln Asd Pro Het Phe Pro Ser Ile Pro Leu Ile 1081 AGC TTC AAG GGC GIT CCA GAT TAT GGA TAC GGA TGT AGA CCA GGA ACG ACG TCA AAG GAC 1140 1361 Ser Phe Lys Gly Val Pro Asd Tyr Gly Tyr Gly Cys Arg Pro Gly Thr Thr Ser Lys Asd 1800 1141 GGT AAT CCT GIT AGT GAC ATG GGG GTA TAT CCC AAA GGC ATG TAC GAC TCT ATA 1200 1381 Gly Asn Pro Val Ser Asd Ile Gly Trp Glu Val Tyr Pro Lys Gly Het Tyr Asd Ser Ile 400 1401 Val Ala Ala Asn Glu Tyr Gly Val Pro Val Tyr Val Thr Glu Asn Gly Ile Ala Asd Ser 420 1261 AAA GAT GTA TTA NGC GGG TG TAC GAC GAA AAC GGA ATA GCA GAT TCA 1260 1421 AAA GAT GTA TTA NGC GGA TGT TAC GAC GAA AAC GGA ATA GCA GAT TCA 1260 1421 AAA GAT GTA TTA NGC GGA TATA TAC GTA ACA GAA AAC GGA ATA GCA GAT TCA 1260 1421 AAA GAT GTA TTA NGC GGA TATA TAC GTA ACA GAA AAC GGA ATA GCA GAT TCA 1260 1421 AAA GAT GTA TTA NGC GGA TATA TAC GTA ACA GAA AAC GGA ATA GCA GAT TCA 1260 1421 AAA GAT GTA TTA NGC GGA TTA TAC GTA ACA GAA AAC GGA ATA GCA GAT TCA 1260 1421 AAA GAT GTA TTA NGC GGA TTA TAC GTA ACA GAA AAC GGA ATA GCA GAT TCA 1260 1421 AAA GAT GTA TTA NGC GGA TTA TAC GTA ACA GAA AAC GGA ATA GCA GAT TCA 1260 1421 AAA GAT GTA TTA NGC GGA TTA TAC GTA ACA GAA AAC GGA ATA GCA GAT TCA 1260 1421 AAA GAT GTA TTA NGC GGA TTA TAC GTA ACA GAA AAC GGA ATA GCA GAT TCA 1260 1421 AAA GAT GTA TTA NGC GGA TTA TAC GTA ACA GAA AAC GGA ATA GCA GAT TCA 1260 1421 AAA GAT GTA TTA NGC GGA TTA TAC GTA TAC GTA ACA GAA AAC GGA ATA GCA GAT TCA 1260 1421 AAA GAT GTA TTA NGC GGA TTA TAC GTA TAC GTA ACA GAA AAC GGA ATA GCA GAT TCA 1260 1421 AAA GAT GTA TTA NGC GGA TTA TA NGC GGA TTA TAC GTA ACA GAA AAC	
961 GAC GGA GAG ACA TIT GIT TAC CIT CCA TAT TIA AAG GGC AAT GAT TGG CTG GGA GTG AAT 1020 121 ASP Gly Glu Thr Phe Val Tyr Leu Pro Tyr Leu Lys Gly Asn Asp Trp Leu Gly Val Asn 140 1021 TAT TAT ACA AGA GAA GTC GIT AAA TAC CAA GAT CCC ATG TIT CCA AGT ATC CCT CTC ATA 1080 141 Tyr Tyr Thr Arg Glu Val Val Lys Tyr Gln Asp Pro Het Phe Pro Ser Ile Pro Leu Ile 360 1081 AGC TTC AAG GGC GTT CCA GAT TAT GGA TAC GGA TGT AGA CCA GGA ACG ACG TCA AAG GAC 1140 161 Ser Phe Lys Gly Val Pro Asp Tyr Gly Tyr Gly Cys Arg Pro Gly Thr Thr Ser Lys Asp 380 1141 GGT AAT CCT GTT AGT GAC ATT GGA TGG GAG GTA TAT CCC AAA GGC ATG TAC GAC TCT ATA 1200 1201 GTA GCT GCC AAT CLA TIT GIA TGG GAG GTA TAT CCC AAA GGC ATG TAC GAC TCT ATA 1200	

Figure 7a

1121 441 1381	C1n UV	AAT Asn	CIA	TAT	CAC Asp	CTC	AGA Ara	GIV	TAC	TTA	CAC	TCC	GCA	TTA	ACC	GAT	AAT	TAC	GAA	TCC	1 tao
1381	CCC	TTA	CCC											Den	inr	VED	Asn	Tyr	Clu	Trp	460
461 1441	717	ren	GIA	Phe	Arg	Het	Arg	Phe	Gly	Leu	Tyr	Clu	Val	AAC ABD	TTC Leu	ATA 11e	ACC Thr	AAA Lys	GAG Glu	AGA Ara	1440 480
481	Lys	Pro	Arg	Lys	Lys	ACT Ser	GTA Val	AGA	CTA Val	TTC Phe	AGA	GAG	ATA	CIT	ATT	AAT	AAT	CCC	CTA	ACA Arg	1500
1501	AGC Ser	λAC	ATC	ACC		CLC								36	116	Asn	Asn	Gly	Leu	Thr	500

Figure 7b(Continued)

PYROCOCCUS FURIOSUS GLYCOSIDASE - 7G1 COMULETE GENE SEQUENCE - 10/95

	, ,										,		. 31	LUUI.	MCX	- 1	0/9	5			•					
	1 .	AFG	II	CC	TC	AA A	VA.G	TTC	CT"	T TO	ic c											CN CN				
		ie t	Phe	e Pr	O G	lu [-Y5	Phe	Lei	. Tr	2 0) ·	016	60	COL	A TC	CG	ST T	17 (CAC	TTI	C)		·~ ~		_
							•	• •• •			P 0	ı y	val	AL a	1 G1:	n Se	r G	ly P	he (Sln	Phe	61.			·	EO
	1 (AT.	AAA	CT	CA	GG A	CG	744	מידי מ				. : .					-	,		•	AGC	, 46	t C	± ¥′	20
2	2: 2	qe/	Lys	Le	J A	τα A		~.	71	. w	CA	CT A	AAC	ACI	GA:	TG	G TO	ic c	AC 7	201	CT.	AGC Azş				
			-			- y	- 9.	~3N	116	, As	וז פ	hr /	re.	The	' Ası	TE	D T	ם א		.~	014	, ACC	, GA	TA	AG.	120
12	ב וי	CB	3 3 **														•			. ~ ~	441	_ AE 3	3 As	D L	V S	40
4	1 7	hr	Aan	Tla		10 7	~~ (500	CIC	: GT	T AC	3T (KZK	GAT	CTI							AAC Aan				
•					- 01) A	el A	Leu	٠Va.	ı se	er (ijy	Asp	Le	. D-	~ ~		46 6	GG	ATT	AAC	AA	T T	AC .	180
18	1 6	20	~==			_														' Y	TTE	מכא	e A	ח די	v =	60
6	1 6	111	Leu	~		W A	AG (₽XC	CVI	CA	5 A1	7	Ċλ	AGA	AAC							TAC		- '	, -	-
			200	- y 1	. 61	n L	ys)	\3p	His	Gli	u Il	e 2	هد	Arn	Tue			T C	TT A	AT	CCI	TAC	AG	A A'	- 4	240
24	1 6	~~										- •			- Lys	Tel	1 61	y L	eu A	.3n	λla	Tyr	Ar	a 1		
	• -	•	ATA	CAC	IC	CA	د حت	CA	ATA	TTC		2 7	~~	~~ "								- , -	~~	, 11	. =	80
•	. .	тĀ	110	Glu	Tr	p Se	er A	T a	He	Phe	D =			CCA	ACG	YC)	IT	T A:	TG	AT.	CTT	CAT	- 5.			
20								•			- • •		гÞ	PIO	The	Tha	Ph	e Il	eλ	30	V=1	GAT Asp	T	. ^	~	300
.ن د	I T	AI.	AXT	GYY	TC	A TZ	A TA	חב		3 ~ =											1	Asp TTG	A.	. Se	: =	100
10	l T	yr :	Asn.	Glu	Se	r Ts	/- N		~	217	• GA	A G	AT (GTA	λλG	ATC	: AC	C AA		20						
						,	· ~	.3.7	reu	1 - 6	GI	n y	зb ,	Val	Lys	Ile	Th	- 1.v		~L .	VCI	TTG Leu	GAG	; GX	S	360
363	T	CA (CiT	CAC				_												٠,	THE	ren	GIU	ı Gl	u	120
121	ي ا	eu 2	Asp	6);;	T1	- 33		AC /	AAG	AGG	CA.	C C	77	3CC	TAC	TRT						AAC Asn				
				-14	11	e VT	A A	מכ.	Lys	Arg	Gl	u V	al 2	L) a	TVE	Tive	200		A G:	נכ א	NTA	AAC	AGC	CT	5	120
421	3.0		~~												٠,٠	- 7 -	~-	, se	r V	11 :	lle	TAT	Ses	Le	u ·	140
141	· ·	~ ,	10.	AAG	حتز	3 II	TA	AG (77.	ATA	GT	נג ז	T C			~~~		_								140
	~) E =	-7.2	زخت	y 3 <u>:</u>	e L	1 2.7	/=1	11.	V.	1 2.	: :	- 1.7	\sim	CAC	TI	: AC	C C 2	TT (CA	TAT Tyr	766		_	400
40.		_ :									٠ ـ .	•		.eu	~==	2.3	Spe	L.L	r Le	u E		Tvr	TEN	* -		480
48:	CA	<u> </u>	λľ	CCC	ATT	حت ٦	န္ မ	T 1	ree.	C3 C	300									-		Tyr Tee	وخد	TE	٠, .	160
161	R1	3 }	Sp	Pro	Ile	G)	11 2				~~~		G I	TA	ACT	XAI	AAC	: AG	בג ב	C 0		TGG Trp			_	
_							- /.		T. a	CTA	Arg	נאינ	a L	.67	Thr	Asn	Lys	Ar	7 2 -	- C	3		GIT	AAG	- 5	640
542	cc	X A	CA.	ACA.	GTT	י בב	A C 1	· ~	_								•		,		TA	Trp	Val	Хэг	1]	180
191	25	o A	TO :	The	Val	T1.	_ ~	W T	TT	GCA	AAG	T)	∵T G	CC	GCT	TAC	ATA					III Phe				
			•		,		4 6.	.u P	ne .	Ala	Lys	Ty	z A	la,	Ala	TVT	T 3		- 10	- ^	AG :	III	CCY	GAT	: €	500
601	AT:	A G	TG (T K 2								-				- , -		~.	Гy	I L	ys !	Phe SGC :	Gly	Asp	> 2	200
-201	11	. v		~~.	~ .	100	y AG	æ λ	CC '	III	AAT	CA	G C	CT :	ATC	CTC				_				_		
_			/	ترد	net	TI	P Se	: I	hr :	Phe	מכת	GI	u P	- ·	4	V-1	611	GTT	, CX	e c	II (SGC (TAC	CTA	. 6	60
661	GC (~~ ,												,e c	ATT	ATT	VA	. G1	u L	eu (Jy'	Tvr	lev	2	20
221	31.			CAC	ICI	GGC	11	c c	CT (CA	GGG	GT.	T C	~» .								Ela (_	~ 0
	~:		רס :	y=	Ser	رنت	r Ph	e P	20 1	2:0	670	V.	1 .	/	WI.	CCA.	CAG	ec c	CC	A A	AG C	TG (200	175	7	20
721											,	-		eu ,	מם.	220	Clu	Ala	· XI	a L	va I	.012	413	71.	·	
	C 1 1	. 0	NC A	NIG ,	ATA	AAT	. ec	A C	AT (-C-T		~~							•	•		iac s		- 1 -		40
241	Let	. H:	Ls >	iec	Ile.	Asn	וגו	a H			1	GC.	T 17	י די	vec (CAG	ATA	AAG	AA	- 11	רד' כ		~	~ > ~	-	~ ~
											Ten	AL:	a T	yr Þ	にす (-ln	Ile	Lys	Lv	3 D)) A	AC A	W 1	~~~		80
781	XXX	CC	TC	AT :	AAG	GAT												•	-3.	• • •		AC A	.nr	610	Z	60
261	Lys	A.	la A	. ae	LVE	D = 5		. ~	W C	AG	CCI	ပေ	C	VA G	TT C	GI .	ATA	ATT	TAC	٠.,		AC A In I				
					-,-	بإدم	36	ב די	/3 G	iu .	Pro	λl	G G	Lu V	al o	lv	Il.	77	7	. ~	L A	AC A	TT	CCA	84	40
841	CII	GC	T T	AT C															1 7	۸.	in A	sn I	le	C) y	28	90
251	Val	Al	A T	ve 1	0	***	GA:	r co	C Y	AC (GAI	TCC	אג	re e	AT C	TT :	226	CC >				su Y YC C				
				3 + 1	rro	rys	As:	2.5	O Y	an)	qt.	Ser	Lv	A e	an V	2)	7	64.X	الان	r Cry	AA	AC G	AC /	AAC	90	00
901	TTC	7	~ ~								_		-,		-, .		-ys	A. a	YYS	Gl	uА	in A	3p /	Asr.	30	0.0
301	Phe	D.	- 11	AC 1	TO.	CCC	CTG	TI	C T	TC C	EAG.	GCC	AT	A C									-			
	• • • • • • • • • • • • • • • • • • • •		Ф. г.	3 3 . 5	er	Gly	Leu	. Ph	e P	he (53 11	111	73		~ ^	~~ !	-CaA	AAA	CTI	λA	T A	IA G	AC S	TT	96	in.
961	-											<i>_</i> a		€ n.	19 1	ys (21 A	Lys	Leu	As	n I	le G	lu i	he	32	
961 321	OAL	GG	T C	VA A	CG	TTT	ATA	GA	T G	C C		~~~														•
321	AJP	G1	y G	lu T	hr	Phe	Ile	1.	n A	1		TVT	CT.	A A	CC	GC X	VAT.	GAC	TCC	AT.	A C.	S G				
									۳ ۸.	+4 2	ΙĐ	ryr	Le	u I.	/3, G	ly 🕽	תכו	A D D	Tro	Tì	- C1	~	11 /	W.1		20
1021	TAC	TA	C AC	.	GC :	GAA	CT.					_						•			e 0,1	y v	BT W	מבי	34	O
341	Tyr	Ty	Th	12 3			V-1	GT.	, w	G T	AT (CAC	CY.	A CC	:A A:	rc T	17	CCT	703							
		•	- •		- -		-41	∀a ,	ı Ti	l I	yr (Gln	Gl	u Pr	0 M	et P	he.	D	5	V1.(- 55	y Va o Le	rg a	.TC	10	
1081	ACC	TT	בב ד			~	_												Jer	716	e PI	O L	eu l	le	36	0
361	Thr	Ph			I	41	-AA	200	KT P	LI C	GC :	TAT	GC	TO	C 20	a c	CT 4			_						
		- •••	- - y	J ()	* A /	- 1 I	Gln	Gly	Y Ty	r C	ly	ryr	نتم	. Cu	- ^	~~ L		- Nu	ACT	CIC	FTC	3 3 2	1G G	AT	11	40
1141	GAC	۸.							-		-	_		- - y	٠,٠	. y P	20 (∍⊺ A	Thr	Lou	1 5e	Х Д А = Ly	/3 A	30	386	
381		A -		C C	רכ א	/CC	GAC	λTX	CG	A TI	GG r	:Aa			+							•		•		
	, cap	~FÇ	. Pr	o V	al 9	er.	qeA	110	(3)		C	=1	T =	T.W	1 00	A G	من (-GG	ATG	TAC	GA	r Ly	A	TA	120	20
1201							•			•	٠,٠	- L U	ret	Ty	r Pr	:0 G	lu (iy:	Met	TYE	Ā-	T TC P Se	- 1	1 0	400	
401	-17	همت	C.	T CX	AC A	uc ·	TAC			T C	٠, ،							•				p Se G GA				•
701	val :	Clu	Al.	a Hi	T	۰ د ۷۰	70-	Class		,	-^ C	T	TAC	GT	ב אכ	C	AG >	AC (GGA	ATA	c-1	G GA	·		126	
						., -	. 1 .	o i y	VA	1 21	OV	a l	Tyr	· Va	1 Th	r G	lu)	מבו	51 v	71-		- 124	, 10	~/\ -		
																		'	- 4 3	116	~7.	4 A.S	p se	: I	420	,

Figure 8a

1261 421	Lys Asp II	C CTA	AGA CC	TAC	TAC	ATA Ile	GCG Ala	AGC	CAC	ATA	AAG	ÁTG	ATA	CAG	AAG	ccc	777	1320
1321	CAG CAT CO				_						Lys	net	116	CIU	Lys	Ala	Phe	440
	CAG CAT CO				_	•						Inr	4 7 7	A	n -	~ `		1380
																		160
																		1440
481	ATT CCC AG Ile Pro Ar	g Glu	Lys Ser	GTG Val	TCG Ser	ATA Ile	TTC Phe	AGA Ara	GAG	ATA	GTA	GCC	AAT	AAT	GGT	GTT	ACG	1500
1501	ANA MAG AT	T (2) N	C) C					•			V41.	W1.7	ASD	Asn	Gly	Val	Thr	500
501	Lys Lys Il	e Glu	Glu Glu	Leu	Leu	YLà	Gly	End	51									

Figure 8b(Continued)

Bankia gouldi endoglucanase (37071)

	,												
9 18 27 36 45 54 5' ATG AGA ATA CGT TTA GCG ACG CTC GCG CTC TGC GCA GCG CTG AGC CCA GTC ACC													
5		•											
	MAE AND THE GOT AND AND COLOR OF AND COLOR OF AND COLOR	CTC: ACC											
	Met Arg Ile Arg Leu Ala Thr Leu Ala Leu Cys Ala Ala Leu Ser Pro	ULL ACC											
		val Thr											
	63 72 91 90 99												
	TTT GCA GAT AAT GTA ACC CTA GAA AGG GAG	108											
	Phe Ala Asp Asn Val Thr Val Gln Ile Asp Ala Asp Gly Cly Lys Lys	CTC ATC											
	The val din He Asp Ala Asp Gly Cly Lys Lys	Leu Tle											
	9.00	120											
	AGC CGA GCC CTT TAC GGC ATG AAT AAC TCC AAC GCA CAA AGC CTT ACC Ser Arg Ala Leu Tyr Gly Met Agn Agn Sor Agn Ala Cal Agc CTT ACC	162											
	Ser Arg Ala Leu Tyr Gly Met Asn Asn Ser Asn Ala Clar Acc	GAT ACT											
	Ser Arg Ala Leu Tyr Gly Met Asn Asn Ser Asn Ala Glu Ser Leu Thr	Asp Thr											
	171 180 180												
	CAC TGG CAG CGT TTT CGC GAT GCA GGT GTG CGC ATG CTG CGG GAA AAT ASP TTP Gln Arg Phe Arg Asp Ala Gly Val Arg Mon Tree CGG GAA AAT	216											
	ARD TED GIR AND DIE COC GAT GCA GGT GTG CGC ATG CTG CGG GAA AAT	GCC CCC											
	Asp Trp Gin Arg Phe Arg Asp Ala Gly Val Arg Met Leu Arg Glu Asn	Cly Cly											
		GTA CTA											
	225 234 243 252 261												
	AND AND AGE ACC AXX TAT AND THE CAR THE CAR	270											
	Asn Asn Ser Thr Lys Tyr Asn Trp Glp Low Mile .	GAT TGG											
	Asn Asn Ser Thr Lys Tyr Asn Trp Gln Len His Len Ser Ser His Pro	متن خور											
	776	_											
		324											
	TAC AAC AAT GTC TAC GCC GGC AAC AAC AAC TGG GAC AAC CGG GTA GCC. TYT Aan Aan Val Tyt Ala Gly Aan Aan Aan Tag GAC AAC CGG GTA GCC.	מישוע בינה											
	Tyr Asn Asn Val Tyr Ala Gly Asn Asn Asn Trp Asp Asn Arg Val Ala	Lau XI											
		reg ITS											
	333 342 351 360 369												
	CAG GAA AAC CTG CCC CCC CAG ACC ACC ACC	378											
	Gln Glu Asn Leu Pro Gly Ala Asp Thr Met Trp Ala Phe Gln Leu Ile	et aag											
	and the lip Ala Phe Gln Leu Ile (ly Lys											
	387 396 405												
		432											
	GTC GCG GCG ACT TCT GCC TAC AAC TTT AAC GAT TGG GAA TTC AAC CAG T	TG CAA											
	Val Ala The Ser Ala Tyr Asa Phe Asa Asp Try Glu Phe Asa Gla S	20 CM											
		er CIU											
	441 450 459 468 477												
	TOG ACC GGC GTC GCT CAG AAM CTC CCC CCC	486											
	Trp Trp Thr Gly Val Ala Gln Asn Leu Ala Gly Gly Gly Glu Pro Asn L	TG GAC											
	and the sty Gly Gly Glu Pro Asn L	еи Азр											
	495 504 513 572 573												
	GGC GGC GGA GCG CTG GTT GAA GGA GAC CCC AAT CTC TAC CTC ATG G	540											
	GIV GIV GIV GIV AT THE STATE GAR GAC CCC AAT CTC TAC CTC ATG G	AT TGG											
	Gly Gly Glu Ala Leu Val Glu Cly Asp Pro Asn Leu Tyr Leu Met A	an Tro											
	·	- L											
	549 558 567 576 585	. 504											
	TCG CCA GCC GAC ACT GTG GGT ATT CTC GAC CAC TGG TTT GGC GTA AAC GC Ser Pro Ala Asp Thr Val Gly Tle Lou App No. 700	334											
	Ser Pro Ala Asp Thr Val Gly Ile Leu Asp His Tro Pho Clause	æ crc											
	Ser Pro Ala Asp Thr Val Gly Ile Leu Asp His Trp Phe Gly Val Asn G	Y Lou											
	603 612 621												
	GCC GTO CGG CGT CGC AAA GCC AAA TAC TGG AGT ATG GAT AAC GAG CCC CC GLy Val Arg Arg Gly Lve Ala Lve Tvr Tro Cor Vice Co	648											
	Gly Val Arg Arg Cly Line Line Arg TAC TGG AGT ATG GAT AAC GAG CCC CC	C ATC											
	Gly Val Arg Arg Gly Lyr Ala Lyr Tyr Trp Ser Net Asp Asn Glu Pro Gl	v T)-											
	657 666 675 684 593												
_	SOUTH GC ACE CAC GAR GRA GRA GRA	702											
-	FIFD Val Gly Thr His Asp Asp Val Val Los Glo Clo Clo Clo Clo Clo Clo Clo Clo Clo C	T TIC											
	Trp Val Gly Thr His Asp Asp Val Val Lys Glu Gln Thr Pro Val Glu As	p Phe											

Figure %

Bankia gouldi andoglucanase (37GP1) (continued)

CTG CAC ACC TAT TTC GAA ACC GCC AAA AAA GCC CGC GCC AAA TTT CCC GGT ATT 720 Leu His Thr Tyr Phe Glu Thr Ala Lys Lys Ala Arg Ala Lys Phe Pro Cly Ile 774 ANA ATC ACC GGT CCC GCT AAT GAG TGG CAG TGG TAT GCC TGG GGC GGT Lys Ile Thr Gly Pro Val Pro Ala Asn Glu Trp Gln Trp Tyr Ala Trp Gly Gly 819 TTC TCG GTA CCC CAG GAA CAA GGG TTT ATG AGC TGG ATG GAG TAT TTC ATC AAG Phe Ser Val Pro Gln Glu Gln Gly Phe Met Ser Trp Met Glu Tyr Phe Ile Lyr 882 CGG GTG TCT GAA GAG CAA CGC GCA AGT GGT GTT CGC CTC CTC GAT GTA CTC GAT 891 Arg Val Scr Glu Glu Gln Arg Ala Ser Gly Val Arg Leu Leu Asp Val Leu Asp 936 CTG CAC TAC TAC CCC GGC GCT TAC AAT GCG GAA GAT ATC GTG CAA TTA CAT CGC Leu His Tyr Tyr Pro Gly Ala Tyr Asn Ala Glu Asp Ile Val Gln Leu His Arg 990 ACG TTC TTC GAC CGC GAC TTT GTT TCA CTG GAT GCC AAC GGG GTG AAA ATG GTA Thr Phe Phe Asp Arg Asp Phe Val Ser Leu Asp Ala Asn Gly Val Low Met Val GAA GGT GGC TGG GAT GAC AGC ATC AAC AAG GAA TAT ATT TTC GGG CGA GTG AAC Glu Gly Gly Trp Asp Asp Ser Ile Asn Lys Glu Tyr Ile Phe Gly Arg Val Asn 1098 CAT TOG CTC GAG GAA TAT ATG GGG CCA GAC CAT GGT GTA ACC CTG GGC TTA ACC Asp Trp Leu Glu Glu Tyr Met Gly Pro Asp His Gly Val Thr Leu Gly Leu Thr 1143 1152 GAA ATG TGC GTG CGC AAT GTG AAT CCG ATG ACT ACC GCC ATC TGG TAT GCC TCC Glu Met Cys Val Arg Asn Val Asn Pro Met Thr Thr Ala Ile Trp Tyr Ala Ser 1197 1206 ATG CTC GGC ACC TTC GCG GAT AAC GGC GTC GAA ATA TTC ACC CCA TGG TGC TGG Met Leu Gly Thr Phe Ala Asp Asn Gly Val Glu Ile Phe Thr Pro Trp Cys Trp 1260 ANC ACC GGA ATG TGG GAA ACA CTC CAC CTC TTC AGC CGC TAC AAC AAA CCT TAT Asn Thr Gly Met Trp Glu Thr Leu His Leu Phe Ser Arg Tyr Asn Lys Pro Tyr 1314 CGG GTC GCC TCC AGC TCC AGT CTT CAA GAG TTT GTC AGC GCC TAC AGC TCC ATT 1323 Arg Val Ala Ser Ser Ser Ser Leu Glu Glu Phe Val Ser Ala Tyr Ser Ser Ile 1368 AMC GAA GCA GAA GAC GCC ATG ACG GTA CTT CTG GTG AAT CGT TCC ACT AGC GAG Asn Glu Ala Glu Asp Ala Met Thr Val Leu Leu Val Asn Arg Ser Thr Ser Glu

Figure 9b(Continued)

Bankia gouldi endoglucanase (37GP1) (continued)

1413 1422 1431 1440 1449 1458
ACC CAC ACC GCC ACT GTC GCT ATC GAC GAT TTC CCA CTG GAT GGC CCC TAC CGC
Thr His Thr Ala Thr Val Ala Ile Asp Asp Phe Pro Leu Asp Gly Pro Tyr Arg

1467 1476 1485 1494 1503 1512
ACC CTG CGC TTA CAC AAC CTG CCG GGG GAG GAA ACC TTC GTA TCT CAC CGA GAC
Thr Leu Arg Leu His Asn Leu Pro Gly Glu Glu Thr Phe Val Ser His Arg Asp

1521 1530 1539 1548 1557 1566
AAC GCC CTG GAA AAA GGT ACA GTG CGC GCC AGC GAC AAT ACG GTA ACA CTG GAG
AEN Ala Leu Glu Lys Gly Thr Val Arg Ala Ser Aep Aen Thr Val Thr Leu Glu

1575 1584 1593 1602 1611
TTG CCC CCT CTG TCC GTT ACT GCA ATA TTG CTC AAG GCC CGG CCC TAA 3*
Leu Pro Pro Leu Ser Val Thr Ala Ile Leu Leu Lys Ala Arg Pro ***

Figure 94 (Continued)

Thermologa maritima Alpha-qalactusidase Complete Gene Sequence (1 c + 3)

• • • • • • • • • • • • • • • • • • • •
5' CTC ATC TCT CTC CAA AWA TVD CTC 100 100 100 100 100 100 100 100 100 10
5' CTG ATC TGT GTG GAA ATA TIC GGA ANG ACC TTC AGA GAG GGA AGA TTC GTT CT
Val Ile Cys Val Glu Ile Phe Gly Lys Thr Phe Arg Glu Gly Arg Phe Val Le
ANA GAG ANA AND 72 81 90
Lys Glu Lys Asp Phys. 72 81 90 99 106 Lys Glu Lys Asp Phys. 72 107 108 108 108 108 108 108 108
Lys Glu Lys Asn Phe Thr Val Glu Phe Ala Val Glu Lys Ile His Leu Gly Trp
117 126 135
ATC TCC GGC AGG GTG AAG GGA AGT CCG GGA AGG CTT GAG GTT GTG
Lys Ile Ser Gly Arg Val Lys Gly Ser Pro Gly Arg Leu Glu Val Leu Arg Thr
ANA GCA CCG GAA ANG GTA CTT GTG ANC ANC TOG CAG TCC TGG GGA CCG TGC AGG
Lys Ala Pro Glu Lys Val Leu Val Am Land
Lys Ala Pro Glu Lys Val Leu Val Asn Asn Trp Gln Ser Trp Gly Pro Cys Arg
GIG GIC GAT GCC TIT TCT TIC AAA CCA CCT GAA ATA GAT CCG AAC TGG AGA TAC
Val Val ASD Ala Pho Son Pho
Val Val Asp Ala Phe Ser Phe Lys Pro Pro Clu Ile Asp Pro Asm Trp Ary Tyr
ACC GCT TCG GTG GTG CCC GAT GTA CTT GAA AGG AAC CTC CAG AGC GAC TAT TTC
The Ala Ser Val Val Come and Asset Che Che Asset Che The Tite
Thr Ala Ser Val Val Pro Asp Val Leu Glu Ary Asm Leu Gln Ser Asp Tyr Phe
233 740
GTG GCT GAA CAA GGA AAA GTG TAC GGT TIT CTG AGT TCG AAA ATC GCA CAT CCT
Val Ala Glu Glu Gly Lys Val Tyr Gly Phe Leu Ser Ser Lys Ile Ala Ris Pro
THE THE GET GIG GAA GAT GGG GAA CIT GIG GCA TAC CITC GAA TAT THE GAT GIC
Phe Phe Ala Val Glu Asp Gly Glu Leu Val Ala Tyr Leu Glu Tyr Phe Asp Val
GAG TTC GAC GAC TTT GTT CCT CTT GAA CCT CTC GTT GTA CTC GAG GAT CCC AAC
Glu Phe Amp Amp Phe Val Pro Leu Glu Pro Leu Val Val Leu Glu Amp Pro Am
447 544
ACA CCC CIT CTT CTG GAG AAA TAC GCG GAA CTC GTC GGA ATG GAA AAC AAC GCG
The Pro Leu Leu Clu Lys Tyr Ala Clu Leu Val Cly Met Glu Asn Ala
399
AGA GIT CCA ANA CAC ACA CCC ACT CCA TCC TCC ACC TCC TAC CAT TAC TTC CTT
Arg Val Pro Lys His The Pro The Gly Trp Cys Ser Trp Tyr His Tyr Phe Leu
, and the true the ten

Figure 10a

Thermotoga maritima Alpha-galactosidane Complete Gene Sequence (2,04,3)

G.	T (~	603			6	12			62	1			63	o .		6.	39			641
			~	. 7(3	. C	c	AG /	ις ς .	CK	: W	C Y	VC	C10	: XV	c CI	$c \propto$	C Y	NG 3	ŊΤ	TTC	641
γs	PL	eu '	Thr	II	o G	Lu G	lu 1	ħr	Lev	ı Ly	3 A	รก	Leu	Ly	s Le	<u>ч у</u>	a L	· -		Db.	Pro
			657						•		_										
TT	CG	AG (370	TIC	: C	NG AS	ra G	ac	CAC	: 60	כ ב זו	ic i	CYY	684 AAC	i G CA	е жт	65 N GC	33			702
Ph	- '	 lu 1	 V=1	Dh.								-								100	CIC
		-		2130		31 13		зÞ	ASP	, VT	a T)	~ ·	Glu	Lys	: As	o Il	e G1	уA	æ	Trp	Leu
אנט	: »/		711	CC)	~	72	0			72	9			738	ı		74	7		•	756
																					756 ርጓአ
Val	l Tr	r »	πā	Gly	٨s	p Ph	e P	50	Ser	۷a.l	G1	น (Glu	Met	גג	Ly	 в Va	1 1	le i	λla	Glu
		7	65			77	4			787	ı			707							
እእር	: œ	TI	TC	λTC	00	G 000	Č XI	ľÀ	TGG	ACC	. cc	c c	T	110	AGI	GIT	.08 TC	r C)	u ,	æ	810 TCC
						J 01	, 1,			****	N.	a .	ID	rne	Ser	vau	. Se	c G1	.u .	DE	Ser
GAT	. CI		19 TC	ልአሮ	GA 2	B2		Y3 /	C	837				846			85	5	•		864
													_								
vsb	Va.	1 P	he .	γ 2 υ	Gli	Hi	s Pr	0	уzb	IIP	Va.	l v	الم	Lys	Glu	λ=0	GI)	r Gl	u P	ΣO	Lys
			73			883				891				900			909)			918
X10	GC.	r 72	AC .	AGX	AAC	TO	λλ	ر ، ع -	w	YYC	λ27	T	AC (œ	CIC	GAI	CIL	. 10	Gλ	XX :	GAT
Met	אנג	1)	y i i	yza	λέπ	TE	. XE	n I	.yz	Lys	Ile	T	yr i	Ala	Leu	λsp	Leu	Se	 r L	~- ·	٨٨٦
			27			936				945				954		_					
CAG	CT			MAC	TCG	CIT	, II	c 0	at :	CIC	TTC	: 17	<u>بر</u>	CT	CIG	J GA	963 AAG	λΤ	s G	cc :	972 DAC
												•									
Glu				I	ىزىد	Dec	PD	- ^	ر چې	tea	Pne	. 50	er :	ær	Leu	VLA	Lys	Met	E G	ly :	ŊΞ
AGG	TAC	98 77		200	D ~~~	990				999	~~		_10	800			1017			10	126
YCC			-																		
Arg	ŢУI	Pb	e I	'nΣ	Ile	Уsb	Phe	e L	en 1	Phe	Ala	G]	y A	IJa '	Val	Pro	Gly	Glu	נג ו	r g I	.ys
		103			:	1044			10	253			10	62		1	1071			10	80
λλG	AAC	AT	אָ ג	בא (CCA	ATT	CXC	C	cc 2	TC	AGA	AA	A G	CC.	ATT	CAC_	ACG	ATC	. Ac	X X	AA
Lys	ķsn	11	e T	hr :	Pro	Ile	Gla	A	la F	he .	λrσ	Ly	 3 G	lv	Ile	Glu	Thr	Tle		·	· ·
	•	108									_	•							~	.9 2	
GCC				ΛA (I Sat	1098 TCT	TTC	. A.	11 TC C	107	CCA	T C:	11	16 CC '	· ·	~~ ¹	.125	~	~	_11	34
								-													
Ma	AUT	GI	y G	in)	ds,	Ser	Phe	1.	le L	eù (Gly	СУ	5 G	ly S	Ser	Pro	Leu	Leu	Pr	O Y	la
		114			1	152			11	61			11	70		1	179			11	98
cic (œν	TCC	. c	דכ כ	SVC	∞	ATG	AC	3G Y	TA (3GA	cc	TC	VC >	וכדו	ccc	CCC	TIC	10	CO	CA
Val	gly	Cy	s V	1 (n	ح- وي	Cly	Hat	Αz	g I	le (Sly	Pro	o λ:	 op 1	dir i	ALa	Pro	Phe	77	 D G	 lv
											-										-,

Figure 10 (Continued)

Thermotoga maritima Alpha-dalactosidade Complete Gene Sequenca (% 3.1.5)

1197 1206 1215 1224 1233 1242 GAA CAT ATA GAA GAC AAC CCA CCT CCT CCT CCT ACC
CAN CAT ATA GAA GAC AAC CCA CCT CCC CCT GCA ACA TOG CCG CTG AGA AAC CC
Glu His Ile Glu Arn Arn Cluster
Glu His Ile Glu Asp Asn Cly Ala Pro Ala Ala Arg Trp Ala Leu Arg Asn Al
The same and the same can can be and the same can be and the same can be and the same can be an area.
Ile Thr Arg Tyr Phe Hat His Asp Arg Phe Trp Leu Asn Asp Pro Asp Cys Leu
ATA CTG AGA GAG GAG AAA ACG GAT CTC ACA CAG AAG GAA AAG GAG CTC TAC TCG
Ile Leu Arg Glu Glu Lys Thr Asp Leu Thr Gln Lys Glu Lys Glu Leu Tyr Ser
1350
TAC ACG TOT GGA GTG CTC GAC AAC ATG ATC ATA GAA AGC GAT GAT CTC TCG CTC
Tyr Thr Cys Cly Val Leu Asp Asn Mer Ile Ile Glu Ser Asp Asp Leu Ser Leu
GTC AGA GAT CAT GGA ANA ANG GTT CTG ANA GAA AGG CTC GGA CTC GGT GGA
Val Arg Asp His Gly Lys Lys Val Leu Lys Glu Thr Leu Glu Leu Leu Gly Gly
THE AND THE GAS GAT CTG AGA TAC GAS AND COMME
Arg Pro Arg Val Gln Asn Ile Met Ser Glu Asp Leu Arg Tyr Glu Ile Val Ser
TCT GGC ACT CTC TCA CCA AAC GTC AAG ATC GTG GTC GAT CTG AAC AGC AGA GAG
Ser Gly Thr Leu Ser Gly Acr Wal
Ser Gly Thr Leu Ser Gly Asn Val Lys Ile Val Val Nep Lin New Car Glu
TAC CAC CTG GAA AAA GAA GGA AAG TCC TCC CTG AAA AAA AGA GTC GTC AAA AGA
Tyr His Leu Glu Lys Glu Gly Lys Ser Ser Leu Lys Lys Arg Val Val Lys Arg
THE TAC TIC TAC GAA CAG CCT CAG ACA CAR TO
Glu Asp Gly Arg Asn Phe Tyr Phe Tyr Glu Glu Gly Glu Arg Glu

Figure 10c(Continued)

Thermotoga maritima β-mannanase (saper) (66P2)

			9			18			27			36			45			54
5,	ATG	GGG	ATT	GGT	GGC		GAC	TCC	TGG	AGC	CCG	TCA	GTA	TCG	CCG	Gλλ	TTC	
	Met	Gly	Ile	Gly	Gly	λsp	qzA	Ser	Trp	Ser	Pro	Ser	Val	Ser	λla	Glu	Phe	Leu
			63			72			81			90			99			
	TT 2	تكلمك		CTT	CAC.	. –	ጥርጥ	777		כדור	بلملمك		AGT	CAC			CTV:	108
	Leu	Leu	Ile	Val	Glu	Leu	Ser	Phe	Val	Leu	Phe	Ala	Ser	Asp	Glu	Phe	Val	Lys
				,														
			117			126			135			144			153			162
	GTG	GAA	AAC	GGA	AAA	TIC	GCT	CIG	AAC	GGA	AAA	GAA	TTC	AGA	TTC	ATT	GGA	AGC
	Val		Agn	GIV	Lvs	Phe	Ala	I.eu	Asn	Glv	Lvs	Glu	Phe	Ara	Phe	Tle	GIV	Ser
	141	010	754		_,_					3	-3-						,	
			171			180			189			198			207			216
	λλC	AAC	TAC	TAC	λТG	CYC	TAC	AAG	YCC	YYC	GGA	λTG	ATA	GAC	AGT	CTT	CIG	GAG
			~~~	~~~	Yor		~~~	Tare	Sar	752	614	WAT	Ile	) en	Ser	Val	Len	Glu
	.ASII	ASII	TYL	TAT.	, aec	nis	171	בעם	261	7011	GLY			740	241	741	26.0	<b>U</b> 1U
			225		•	234			243			252			261			270
	AGT	GCC	YCY	GAC	ATG	CCT	AΤλ	λλG	GTC	CIC	λGλ	ATC	TGG	CCI	TTC	CIC	CAC	CCC
													~					
	Ser	Ala	Arg	ASP	met	GIY	me	Γλa	VAI	ben	AFG	116	TIP	GIY	Pne	Leu	ASP	GJA
			279			288			297			306			315			324
	GAG	AGT	TAC	TGC	λGλ	GAC	AAG	AAC	ACC	TAC	ATG	CAT	CCI	GAG	ccc	GGT	GTT	TIC
	Glu	Ser	TYX	CAR	yığ	yab	Lys	Asn	Thr	TYI	Met	Hıs	Pro	Glu	Pro	Gly	Val	Phe
			333	1		342			351			360	,		369	,		378
	GGG	GTG			GGA			AAC			AGC			GAA			GAC	TAC
	Gly	Val	Pro	) Glu	Gly	Ile	Ser	Asn	Ala	Gln	Ser	GJA	. bpe	Glu	·Arg	Leu	yab	Tyr
			387	,		396			405			414			423			432
	ACA	GTT			GCG			CTC	_		λλλ			λTI			GTG	AAC
														· '				
	Thr	Va]	Ala	Lys	, YJa	Lys	Glu	Lev	Gly	Ile	: Lys	Let	ı Val	. Ile	· Val	Leu	Val	Asn
						450			459			466	3		477	,		486
	AAC	. 4rCc	441 3 GAC		. TTC			ATC			TAC			TGG	-		GGA	ACC
	Asn	TE	ak c	y yaz	Phe	: Gly	, G12	r Het	. Asr	Glr	TYX	· Val	LArg	Tr	Pho	e Gly	. cja	Thr
				_								500						E 4 0
	C) a		49!			504 TNC		. Gan	513 CAC		2 ATY	522 344 -		. G.M	53: TA		חמג	540 TAC
	CA1	. LA	_ UA(												- '		, ~~~	
	Ris	Hi:	B AS	b yai	p Phe	יעד פ	: Arg	J Ası	Gli	Ly:	Ile	e Ly	s Glu	ı Glı	ı Ty:	r Lys	Lys	Tyr
				_				-							-			

Figure 11a

Thermotoga				ga	mar:	itim	a B	β-mannanase (								. /		.a .s
											/-000	-	( = (	onti	Due	n (	S (5 1	لمرا
CTV	- m-	549			55	В		56	7		576	5		585	5		594	
	100		r CTC	GT.	y yy	CA:	CIC	C AA	T AC	CTAC	) ACC	GG)	GT	י ככז	TAC	: AGC	594 GAA	
Va:	Se	Phe	Lei	l Va														
					- 75,	ı nıı	ya.	LAS	n Thi	ר. יו	The	: G1)	' Val	Pro	Typ	Arg	Glu	
		603	3		61:	2 .		62	1		630	,		639				
GAC	CCC	) ACC	: ATC	: ATY	GCC	TC	GAC	CT	T GC	A AAC	: GY	, cc	ccc	ביאם . גרם	, , ,		648 GAC	
Cl																ACC	GAC	
GIC	Pro	3.17.	Ile	: Met	· Ale	Tr	Glu	l Lei	u Ala	Asn	Glu	Pro	λrg	Cys	Glu	Thr	λsp	
		657			666									_		• • • • •	, asp	
AAJ	TC			. ACC	יייי כייי		CAC	67! TC	· ~~		684			693	i		702 እእር	
										AAG	GAG	ATG	AGC	TCC	TAC	<b>ATA</b>	AAG	
Lys	Ser	Gly	, γ γ	The	Lev	Val	. Glu	TI	val	Lvs	Glu	Met					Lys	
								_					Ser	ser	JAI	Ile	Lys	
A CT	. ~	711			720			725	•		738			747			756	
		GAI		AAC	CAC	CIC	GIG	GC1	CIC	ecc	CYC	GAA	CCA	TTC	TTC	AGC	AAC	
Ser	Leu	λsp	Pro	. y e.		100												
						Leu	AGT	YTE	VAI	Gly	λsp	Glu	Gly	Phe	Phe	Ser	λsn	
		765			774			783	l		792			201				
TAC	GYY	CCY	TIC	ΧAA	CCI	TAC	CCI	GGA	GAA	GCC	GAG	TGG	CCC	801	220		810	
TYT	GIR	GIĀ	Phe	Lys	Pro	TYI	Gly	Gly	Glu	λla	Glu	Trp	λla	Tyr	λsn	Glv	Tro	
		819			828									_	_			
TCC	GGT			TGG	AAG	AAG	مبلح	837	***	ATA	846			855			864	
Ser	Gly	Val	yeb	Trp	Lys	Lys	Leu	Leu	Ser	Ile	Glu	Thr	Val				Thr	
		_											•	vah	rne	Gly	The	
<b>TTC</b>	CNC	873			882			891			900			909			918	
			TAT	CCG	TCC	CAC	TGG	CCI	CIC	AGT	CCA	GAG	AAC	TAT	GCC	CAG	TGG	
										Ser								
								0.,	•41	Set	Pro	CIB	Asn	TYI	λla	Gln	TIP	
		927			936			945	•		954			963			022	٠
GGA	GCG	λAG	TGG	ATA	GAA	GAC	CAC	ATA	AAG	ATC	GCA	λλλ	GAG	ATC	GGA	3 Z A	972	
01,	~14	Lys	IIP	TTE	GIU	Asp	His	Ile	Lys	Ile	λla	Lys	Glu	Ile	Gly	ГУз	Pro	
		981			990			000			000							
CTT	GTT	CTG	GAA	GAA	TAT	GGλ	ATT	CCA	AAG	AGT	GCG.	CCX		1017		1	.026	
Val	Val	Leu	Glu	Glu	Tyr	Gly	Ile	Pro	Lys	Ser	λla	Pro	Val	Asp	Arc	Thr	112	
												-			9	****	VTE	
ATC		.035	حبد	شتان [	1044	~>~	1	053		1	062		. 1	071		1	080	
					~~~	OAT	CIG	GTC	TAC	GAT	CIC	GGT	GGA	Gat	GCY	GCG	ATG	
										Asp								
				_		-			- 3 -		-€u	O L Y	GΙΛ	ASD	Glv	Ala	Met	

Figure 11b(Continued)

	The	xmot (00'a	mar		. a									
•			,,		· · ·	β-20	enne:	Dase	(55	(C)	((onti	nu e	a, (é	61
TTC T	10: '4 33	16: Car	C	109	9	11	107		111	.6		172	5		•
TTC T				G GG	NATC	GGG (EVY C	GT TC	G GA	C AG	A GA	C GA	G YC)	, ága	1134 747
Phe T	ED We	et Le	u Al	a Gly	/ Ile	Gly	lu G	ly Se							
	116			3 3 5 6											
TAT C	CC CJ	C TAC	C GA	C GG1	TTC	11 AGA A	.61 .ma .co	~	117	0		1179	•		1188
TAT C								C AA	C CA	C CY	C AG	בים ד	(42)	GCC	GAA
Tyr P	ro As	D LAI	. yai	Gly	Phe	Arg I	le Va	l As	n As	P As	 P Se:	 E Pro			
CTG AT	ra ag	y Cyy	TAC	: GCG	λλG	CTGT	10 YY	C AC	122	4 T C3:1		1233	l .	1	L242
Leu Il	 le Ar	 a Glu	~~~								- GA(λGλ	GAA	CYC
Leu Il		, ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,		VIG	rys	Leu P	he As	n Thu	: G13	/ Glu	ı Ası	Ile	λrg	Glu	λερ
	125	7		1260											
ACC TG		r rrc				. wor ou	بر بر	CATC	GAG	ATC	: AAA	AAG	ACC	تىرى 1	.296
Thr Cy	ra Se:	Phe	Ile	Leu	Pro	LVS As	en Gl								
	130	-				-,	, J G1	y net	GIU	Ile	Lys	Lys	Thr	Val	Clu
CTG AG	C CC.	s regr	GTT	1314 ·	GAC '	132	3 .		1332			1341		1	350
GTG AG						TAC AG	C AAC	λCG	TIT	GYY	AAG	M	Tet	ണും വ	AAA
Val Ar	g Ala	GIY	Val	Phe	yab ;	lyr Se	x Asr	Thr	Phe	Glu	Lvs	Len			
	1750	i .	•	1250											
GTC GAN	A GAI	, CLC	GTT	TTT	GXX 3	137 AT GA	, G Ata	GAG	1386 Cat		CC>	1395		1	404
Val Gli		Len	 V-)								GGA	TAC	GGA .	ATT :	TAC
Val Glu	,	260	·	Pne	GIN Y	sn Gl	u Ile	Glu	His	Leu	CIA	Tyr	Gly	Ile 7	Y I
	1411		•	122											
GGC TTT	GAT	CTC	GAC	ACA .		OG AI	- CCG	GAT	CCY	GλA	CAT	GAA	ATG .	14 רויר כ	-1-r 158
Gly Phe	Asp	Leu	Asp	Thr :	Thr A										
	1457		_					لوحم	GIA	GIR	His	Glu	Met 1	Phe L	eu
GAA GGC	CAC	TTT	CAG:	4 / 6 GGA 1		1485	5. • • • •	1	494		1	503		15	12
GAA GGC						-0 010	, AAA	GAC	TCT	ATC	λλλ	GCG :		TG G	TG
Cla Cla	His	Phe (Sln (Gly [ys Ti	ur Val	Lys	λsp	Ser	Ile	Lvs	 Ala 1		·	- <u>-</u>
	1571		4 4												
AAC GAA	GCX	CGG 1	CAC (er c	ייר כר	'A CAC	GAA	GTT (548 Catr (ا فلمنص	1	557		15	66
Asn Glu	Ala	Arg 1	· ·	 '-								rcr (CA G	AA G	AG
Yau Cln		,	yr (ar D	eu XI	a Glu	Glu	Val .	de/	Phe .	Ser .	Ser E	, ro C	lu G	 Lu
7	575		1 6	D 4											
TG AAA	AAC	1GG 1	GG A	••		~ ~~	100	CAG (CA (SAC ?	TTC (200 4	'CA C	16: רזי פו) C
al Lys	Asn	Trp T	TP A	sn S	 er Gl	 Y Thr	Tr	 Gla :							-
								- 111 F	-TA (ıu j	Phe (ily s	er P	ro As	P

Figure 110(Continued)

		The	Fact	Oga	DA :	ci t i v		s β-mannanase						•				1.	,
	•			+ 7			-	P-22	****	285	•	CEC	23)	(c	ont	inue	a) ((6G	P.2)
AT	T GA	162	29		16	3.0			_									1674 A CTG	•
11	e G1	u Tr	D YE	312 G)	y G	lu Va	1 G1		n G)	 У Л	 la	Leu		 n [.e				A CTG s Leu	
		TON	3		160	2													
ccc	: cc	λ λλ 	G AG	C CX	C TO	G GX	λGλ	y Cl	G YC	ia G	TA	GCY	AG	G AA	171 3 TT	9 C (2)		1728 A CTC	
Pro	G1	y Ly	s Se	r As	p Tr	p G1	 u Gl	 u Va	 1 Ar		 al	 Ala	A					Leu Teu	
		. 1/3	,		374	c													
TCA	GA	A TG	T CA	G AT		C GA	G TA	C CY	TA D	C T	١c	ATT	CC	A AAG	177:	3 CAG	GGZ	1782 CTC	
Ser	Gl	Cy:	s Gl	u Il	e Le	u Gli	יעד י	r Ası	11	e 12	~ <u>-</u>	 Ile	Pro	Ası	 1 Val	Glu		Leu	
		1793	L		180	n		4											
						5 IAC	. GC(s GTT	נ כז(3 YY	C	ccc	GGC	TGG	GTG	AAG	ATA	1836 GGC	
rys	G17	, yrč	Let	υ λτς	Pro	TYI	: Ala	Val	Lev	a As	n l	Pro	Gly	LIP	Val	Lys	 Ile	Gly	-
		1845	;		185/														
Len				•		AAC	. .	· GAA	AGT	. GC		2 X G	ATC	ATC	ACT	TIC	೧	GGA	
	روحہ	net	. ASI	1 Asn	Ala	λsn	Val	Glu	Sei	Al	a (Glu	Ile	Ile	Thr	Phe	Gly	Gly	
AAA	GAG	1899 TAC	λGλ	, YCY	1908 TTC	CAT	GTA	1917	محد ۾		19	26			1935			L944	
Lys	 Glu	Tyr	λra	 \ \ra	Pho	 Vi-							GAC	AGA	ycy	GCG	eee	GTG	
		1953	_	-3	1000	His	Val	Arg	116	Gli	1 P	,pe	Asp	Arg	Thr	Ala	Gly	Val	
AAA	GAA	CIT	CAC	λTλ	GGY TAP5	GTT	GIC	1971 GGT	CAT	CAT	19	80	300		1989		1	998	
Lys 	Glu	Leu	His	-i- Ile	Glv	 Va 1	 Val				-		~	TAC	GAT	GGA	CCG	ATT	
	2	2007			3	Val							Arg	TYE	Asp	Gly	Pro	Ile	
TTC .	ATC	GAT	AAT	ണ്ട	AGA	CTT	72T	2025			20	34		2	043				
 Phe	 Ile				•				AGA 	ACA	. G	GA (GGT 	ATG	TGA	3.			
Phe :		, Lap	wan.	ATT	Arg	Leu	TAX	Lys	λrg	Thr	G:	ly (Slv	Met	• • •				

Figure 11d (Continued)

AEPII la β-mannosidase (63gR1)

5' ATG CTA CCA GAA COC 27 36 45
5' ATG CTA CCA GAA GAG TTC CTA TGG GGC GTT GGG CAG TCA GGC TTT CAG TTC GA
Met Leu Pro Glu Glu Phantage Transport
Met Leu Pro Glu Glu Phe Leu Trp Gly Val Gly Gln Ser Gly Phe Gln Phe Gl
ATG GGC GAC AAG CTC AGG AGG CAC ATC GAT CCA AAT ACC GAC TGG TGG AAG TG
THE AGG AGG CAC ATC GAT CCA AAT ACC GAC TGG TGG AAG TG
Het Gly Asp Lys Leu Arg Arg His Ile Asp Pro Asn Thr Asp Trp Trp Lys Tr
GTT CGC GAT CCT TTC AAC ATA AAA AAG GAG CTT CTC 250 153 16.
GTT CGC GAT CCT TTC AAC ATA AAA AAG GAG CTT GTG AGT GGG GAC CTT CCC GA
Val Arg Asp Pro Phe Asn Ile Lys Lys Glu Leu Val Ser Gly Asp Leu Pro Gl
1/1 100
GAC GGC ATC AAC AAC TAC GAA CTT TTT GAA AAC GAT CAC AAG CTC GCT AAA GGC
Asp Gly Ile Asp Asp Com Color
Asp Gly Ile Asn Asn Tyr Glu Lou Phe Glu Asn Asp His Lys Leu Ala Lys Gly
275
CTT GGA CTC AAC GCA TAC AGG ATT GGA ATA GAG TGG AGC AGA ATC TTT CCC TGG
Leu Gly Leu Asn Ala Tor Are 71 Gr
Leu Gly Leu Asn Ala Tyr Arg Ile Gly Ile Glu Trp Ser Arg Ile Phe Pro Trp
2/9 700
CCG ACG TGG ACG GTC GAT ACC GAG GTC GAG TTC GAC ACT TAC GGT TTA GTA AAG
Pro Thr Trp Thr Val Asp Thr Glu Val Glu Phe Asp Thr Tyr Gly Leu Val Lys
333 342 351 360 369 370
GAC GTT AAG ATA GAC AAG TCC ACC CTT GCT GAA CTC GAC AGG CTG GCC AAC AAG
Asp Val Lys Ile Asp Lys Ser Thr Leu Ala Glu Leu Asp Arg Leu Ala Asn Lys
GAG GAG GTA ATG TAC
GAG GAG GTA ATG TAC TAC AGG CGC GTT ATT CAG CAT TTG AGG GAG CTC GGC TTC
Glu Glu Val Met Tyr Tyr Arg Arg Val Ile Gln His Leu Arg Glu Leu Gly Phe
441 456
AAG GTC TTC GTT AAC CTC AAC CAC TTC ACG CTT CCA ATA TGG CTC CAC GAC CCG
THE CAC TIC ACE CIT CCA ATA TOG CTC CAC GAC CCG
Lys Val Phe Val Asn Leu Asn His Phe Thr Leu Pro Ile Trp Leu His Asp Pro
ATA GTG GCA AGG GAG AAG GCC CTC ACA AND GOO 531 540
THE
Ile Val Ala Arg Glu Lys Ala Leu Thr Asn Asp Arg Ile Gly Trp Val Ser Gln

Figure 120

	-	AEPII la β-mannosidas					da s	# (63GB1)				(continued)						
		549			558	t		567	,		576	•						
AGG	ACA	GTI	CII	, CYC	TIT	ccc	AAG	TAT	CCI	GC1	TAC	λTC	GCC	585	CCC	~~~	594 GGA	
Arg	Thr	· Val	Val	Glu	Phe	λla	Lys	Tyr	λla	Ala	Tyr	Ile	Ala	His	Ala	Leu	Gly	
		603			612			621						1.				
GAC	CTC	GTG	GAC	YCY	TGG	AGC	ACC	TTC	AAC	CAN		ATG	GTA.	639	~~~	c> -	648 CTC	
ASP	Leu	Val	λερ	Thr	Trp	Ser	Thr	Phe	λsn	Glu	Pro	Met	Val	Val	Val	Glu	Leu	
		657			666						684							
GGC	TAĆ			ccc	TAC	TCA	GGA	طمامل ۱۳۵۵	~~	CCC	GGA	~~~		693			702	
Gly	TYE	Leu	λla	Pro	Tyr	Ser	Gly	Phe	Pro	Pro	Gly	Val	Met	λsn	Pro	6111	λla	
						•										914	UTG	
CCC	B 3/C	711		3700	720			729			738			747			756	
						AAC	AIG	ATA	AAC	CCC	CAC	GCC	TIG	CCY	TAT	λλG	ATG	
Ala	Lys	Leu	Ala	Ile	Leu	Asn	Met	Ile	Asn	Ala	His	11-	7.00	33-				
												n.Lu	Dea	VIE	TYT	LYS	Met	
3.003		765			774			783			792			801			810	
ATA	AAG	AGG	TIC	GAC	ACC	AAG	AAG	GCC	CAT	GAG	GAT	AGC	AAG	TCC	CCT	GCG	GAC	
									~~-		уа ъ 		_					
	•					3	<i>-</i> 55 -	714	ىرىم	GTA	vab	ser	ГЛЗ	Ser	Pro	λla	Asp	
	*	819			828			837			846			855			864	
GTT	GGC	ATA	ATT	TAC	AAC	AAC	ATC	GGT	CII	CCC	TAC	CCT	AAA	GAC	CCI	λλC	GAT	
	3			4.3.4	V211	ASII	110	CIA	AT	AIA	Tyr	Pro	Lys	yzb	ΡĘΟ	λsn	yzb	
•		873			882			891			900			909			918	
CCC	AAG	CAC	CIT	λλλ	GCA	CCC	GAA	AAC	GAC	AAC	TAC	TTC	CAC	AGC	GGA	CTG	TTC	
													_					
	Dy S	NS P	VAI	rys	YTZ	ATE	GIU	Asn	yzb	Asn	Tyr	Phe	His	Ser	Gly	Leu	Phe	
		927			936			945			954			963				
TTT	GAT	GCC	YLC.	CAC	λλG	CCT	AAG	CTC	AAC	ATA	GAG	TTC	CAC	GGC	GAA	110	972	
rne	YZĐ	VIG	116	H15	Lys	Gly	Lys	Leu	yzu	Ile	Glu	Phe	Asp	Gly	Glu	λsn	Phe	
		981			990			999		-	1008	,						
GTA	XXX	GTT	λGλ	CAC	CTA	AAA	GGC	AAT	GAC	TGG	ATA	œ	محدد .	1017	mx o		.026	
Val	Lys	Val	YLA	His	Leu	Lys	Cly	naA	Asp	Trp	Ile	Cly	Leu	Asn	Tyr	Tyr	Thr	
		.035			.044			053										
CGC			GTT			TCG	CAG	222	AAG	אנד. י	CCY	AC-T	ן בידה	071		1	080	
Arg	Glu	Val	Val	λrg	Tyr	Ser	Glu	Pro	Lys	Phe	Pro	Ser	Ile	Pro	Leu	Ile	Ser	

Figure 12b(Continued)

p-mannosidase (63981) (continued)

1089
1089 1098 1107 1116 1125 113 TTC AAG GGC GTT CCC AAC TAC GGC TAC TCC TGC AGG CCC GGC ACG ACC TCC GC Phe Lys Gly Val PTC Acc TCC TGC AGG CCC GGC ACG ACC TCC GC
Pho The GGC TAC TCC TGC AGG CCC GGC ACG ACC TCC GC
Phe Lys Gly Val Pro Asn Tyr Gly Tyr Ser Cys Arg Pro Cly Thr Thr Ser Al
1143 1152 115
1143 1152 1161 1170 1179 118
Val Ser Asp Ile Gly Trp Glu Val Tyr Pro Gln Gly Ti-
1197 1206 1215 1224 1233 1242 GAC TCG ATA GTC GAG GCC ACC AAG TAC AGT GTT CCT GTT TAC GTC 1242
GAC TCG ATA GTC GAG GCC ACC AAG TAC AGT GTT CCT GTT TAC GTC ACC GAG AAC
Asp Ser Ile Val Glu Ala Thr Lys Tyr Ser Val Pro Val Tyr Val Thr Glu Asn
loss by Ser Val Pro Val Tyr Val Thr Glu Asn
1251 1260 1269 1278 1287 1796
THE CITY AGG CCA TAC TAC ATA COME
Gly Val Ala Asp Ser Ala Asp Thr Leu Arg Pro Tyr Tyr Ile Val Ser His Val
1305 1314 1323 1332 1341 1350 TCA AAG ATA GAG GAA GCC ATT GAG AAT GGA TAC CCC GTA AAA GGC TAC ATG TAC Ser Lya Lle Gly Gly at
Ser Lys Ile Glu Ala Ile Glu Asn Gly Tyr Pro Val Lys Gly Tyr Met Tyr
1359 1368 1377 1386 1395 1404
The life life life life
Trp Ala Leu Thr Asp Asn Tvr Glu Tro No.
Trp Ala Leu Thr Asp Asn Tyr Glu Trp Ala Leu Gly Phe Ser Met Arg Phe Gly
THE ALL MAN AND AND COME COME AND
Leu Tyr Lys Val Asp Leu Ile Ser Lys Glu Arg Ile Pro Arg Glu Arg Ser Val
1467 1476 1485 1494 1503 1512 GAG ATA TAT CGC AGG ATA GTG CAG TCC AAC GGT GTT CCT AAG GAT ATC AAA GAG
Elu Ile Tyr Arg Arg Ile Val Gln Ser Asn Gly Val Pro Lys Asp Ile Lys Glu
THE CTG ANG GGT GAG GAG ANA TGA 3.
lu Phe Leu Lys Gly Glu Lys ***

Figure 12C(Continued)

OCI/4V Endoglucanase (33Gp)

					9			18															
5.	ATY	GG	Tλ	Gλ	A AC	SA C	AC 7	70	AGJ	L TA	T G	27 Pr (طعلت) ere	:	36 A			4 :	5			54
	Mai					·								~		C A	cc	CIG	TT	L CI	T G	T	ATC
	••••	_ ,	aı	GII	ı Ar	g H	is. P	he	Arg	J.	r Va	ıl I	Сeл	Ile	; c)	s T	hr	Leu	Phe			-	
	CIK	; C;	ſλ	ATC	TC	X IX	C A	CI	CAG	TG:	י כסים	ia a		2 20	9	A C			99	•			108
	Lev													~~.	. GA	A C(CA.	AAC	AAA	AG	A GI	G	TAA
			= 4	116	. >e	∓ 5€	er T	hr	Gln	Cλε	G)	уL	ys	λsn	Gl	 u P:	:0 2	Asn	Lve	·			
	AGC	À	G	Gλλ	CA	G TC	λG	LT.	CCT	GAA	λG	r G	λT	ACC:	14	4 C TC			153			1	62
	Ser		-															-CA	TIT	GY)	TA:	C A	AC
			_	51 u	GII	. se	T V	11 /	Ala	Glu	Se	r A	sp	Ser	λει	 n Se	r)	lla.	Phe	Glu	T.		
				171			7 (•															
	AAA 	λT	G	GTA	GG?	KK 1	A GC	ia (STA	λλτ	AT	r G	SA.	λλτ	CC.	ىلمىكى بۇ چ) ~	• > •	207			2	16
	Lys	Me	_ _ '	 Va 1	612														GCT	CCI	, LIK	: G	AA
	Lys		_			Uy.	2 G1	У	/ai	YSD	Ile	G)	ly .	Asn	λlε	Le	u G	lu	λla	Pro	Phe		 lu
,				225			77																
	GGA	GC.	г :	rgg	GGA	GT	A AG	A A	TT	GλG	GAT	, cs	u :	TAT	TIT	, cy	SΑ	TA .	∠61 ATA	330		2	70
	Gly	Ala	3 :	Trp	Gly	Va.	l Ar	 a T	٦.		\										~~~	. A	
	Gly	•		_	Ī			y -		GIU	AST) G1	u	ŊΥ	Phe	Gl	ıI	le	Ile	Lys	Lys	A	rg
			- 2	279			20	0															
	GGA			WI.	TCT	GIT	' AG	G λ	T	CCC	λTλ	λG	A 2	rec	TCA	GC		AT I	ATA	TCC	GAA	32	24
	C1A	Phe	:)	sp	Ser	Va]	Ar	ı	le i	Pro													-
			,									~_	y .	IP	Per	λla	H:	is :	Ile	Ser	Glu	Ly	' S
	CCA (CCA	د 1 ،	33 'AT	CAT	א ייי	34:	?			351				360			3	369				• •
	CCA		-					- A	3G /	LAT.	TTC	CI	CG	XA.	уСУ	GII	, Y)	vc' c	AT	GTT	CIC	GX	T.
	Pro :	Pro	T	УĪ.	yzb	Ile	λs) Aı	rg)	Lsn	Phe	Lev		1u :		V-1							-
			3	87											~ 9	Val	AS	in E	lis '	Val	Val	λs	P.
ž	AGG (CT	c	LL (SAG	AAT	395	, ,		C	405				414			4	23			43	2
٠	AGG (-							·	OLV.	ATC	: A:	TC I	LAT	ycg	CY	c c	AT	TT.	GAA	GA.	A
	Arg }	lla	L	eu (ilu	yzu	Yau	Le	u T	pr 1	Val	Ile	2 I	le j	Lsn	Thr	H;						-
			44	47			AFA																
C	TC I	TAT	C	u c	:AA	CCG	GAT	λA	λт	אכ מ	159 300	Cht	, ,	4	88		_	4	77			48	6
1															16	GTG	Gλ	AA	LT I	CC 1	AGA	CAC	3
•	eu T	ΥŢ	6	ın (lu	Pro	yab	Ly	5 T	AI G	ly	Asp	Va	ıl L	æu	Val	Gl	u I	 le 7	,			•
			48.5	' '			E 10 4			_													
A	TT G	CA	A.	A T	TC ·	LT	λλλ	GA'	TT	AC C	.00	GAA	AA	5 Tr.C	22 TC:	~~~		5	31			540)
I	 le A) a	 L.:		·												TT	r G	A.A.A	ו אנ	CAC .	AAC	:
-	le A	- 4	۷,	s P	110	rne	Lys	λs	נג פ	/I P	ro	Glu	zΑ	n L	eu .	Phe	Phe	- G:	lu I	 le 7	`~-	 Nan	· 1
																			-		3 - 1		•

Figure 130

and the control of th
OC1/4V Endoglucanase (33GP1) (continued)
549 558 567 (JiGPI) (continued) GAG CCT GCT CAG AAC TOTAL ACT TOTAL CONTINUE ACT TO TOTAL CONTINUE ACT TOTAL
GAG CCT GCT CAG AAC TTG ACA GCT GAA AAA TGG AAC GCA CTT TAT CCA AAA GTG
THE THE TAKE GCA CIT TAT CCA AAA GTG
Glu Pro Ala Gln Asn Leu Thr Ala Glu Lys Trp Asn Ala Leu Tyr Pro Lys Val
TYP Pro Lys Val
603 612 621
CIC AAA GTT ATC AGG GAG AGC AAT CCA ACC CCC 630 639 648
CTC AAA GTT ATC AGG GAG AGC AAT CCA ACC CGG ATT GTC ATT ATC GAT GCT CCA Leu Lys Val Ile ATC GAT GCT CCA
Leu Lys Val Ile Arg Glu Ser Asn Pro Thr Arg Ile Val The
The Val Tie Ile Asp Ala Den
880 700 000 666 675
ANC TAT AGC GCA GTG AGA AGT CTA ANA TOTAL 693 702
AAC TGG GCA CAC TAT AGC GCA GTG AGA AGT CTA AAA TTA GTC AAC GAC AAA CGC
Asn Trp Ala His Tyr Ser Ala Val Arg Ser Leu Lys Leu Val Asn Asp Lys Arg
711 720 292 Deu Val Asn Asp Lys Arg
ATC ATT GTT TCC TTC CAR THE 729 738 747
ATC ATT GTT TCC TTC CAT TAC TAC GAA CCT TTC AAA TTC ACA CAT CAG GGT GCC
Ile Ile Val Ser Phe His TVT TOT Clu Des
Ile Ile Val Ser Phe His Tyr Tyr Glu Pro Phe Lys Phe Thr His Gln Gly Ala
765 774 775
GAA TGG GTT AAT CGG AGG 783 792 801
GAA TGG GTT AAT CCC ATC CCA CCT GTT AGG GTT AAG TGG AAT GGC GAG GAA TGG
Glu Trp Val Asn Pro Ile Pro Pro Val Arg Val Lys Trp Asn Gly Glu Glu Trp
The Pro Pro Val Arg Val Lys Trp Asn Gly Gly Cly Co
GAA ATT AAC CAA ATC ACA ACC ACC GAS 646 855
THE AND THE AND THE GOT GAC TOG GEN AND CAN
Glu Ile Asn Gln Ile Arg Ser His Phe Lys Tyr Val Ser Asp Trp Ala Lys Gln
The bys Tyr Val Ser Asp Trp Ala Lys Gln
AAT AAC GTA CCA ATC TTT CTT GGT GAA TTC GGT GGT GAA TTC GGT GGT GAA TTC GGT GA
AAT AAC GTA CCA ATC TTT CTT GGT GAA TTC GGT GCT TAT TCA AAA GCA GAC ATG
Asn Asn Val Pro Ile Phe Leu Gly Glu Phe Gly Ala Tyr Ser Lys Ala Asp Met
927 Pac
GAC TCA ACC CTT 100 945 954 060
GAC TCA AGG GTT AAG TGG ACC GAA AGT GTG AGA AAA ATG GCG GAA GAA TTT GGA
ASD SET ATT VALUE OF THE SET OF T
Asp Ser Arg Val Lys Trp Thr Glu Ser Val Arg Lys Met Ala Glu Glu Phe Gly
981 990 and
TTT TCA TAC GCG TAT TGG GAA TTT TGT GCA GGA TTT GGC ATA TAC GAT AGA TGG
Phe Ser Tor 110 GAA TIT TGT GCA GGA TTT GGC ATA TAC GAT AGA TGG
Phe Ser Tyr Ala Tyr Trp Clu phe C
Phe Ser Tyr Ala Tyr Trp Glu Phe Cys Ala Gly Phe Gly Ile Tyr Asp Arg Trp
1035 1044 1053
TCT CAA AAC TOG ATC GAA CCA TTG GCA ACA GCT GTG GTT GGC ACA GGC AAA GAG Ser Gin And Togan Communication of the com
THE SEA ACA GCT GTG GTT GGC ACA GGC AAA GAG
Ser Gln Asn Trp Ile Glu Pro Leu Ala Thr Ala Val Val
Ser Gln Asn Trp Ile Glu Pro Leu Ala Thr Ala Val Val Gly Thr Gly Lys Glu
TAR 31

Figure 136(Continued)

Thermotoga maritima Pullulanase (6023)

				9																
5.	ATG	GA	тс	TT	ACA	AA	عت : T	8			27			3	6			15		54 CA AAJ
										IC A	ra (3TG	yc	2 CIV	S AA	CA	G 70	3 5 C	AG G	CA AA
	Met	λs	p L	eu	Thr	Lys	. Va	1 G)	y II	le T)	 /- 1								
									•			- 41	N. C	Let	iak i	3 G1	u Tz	D C	ln A	 la Lys
				F.Z			~	•					•							
				CA.	АЛЛ	GAC	: AG	GTI	C YI	y C	IG A	XT/	λλλ	GAC	. GG/	L AAG		'Y.	~	10s
	λsp	Va.	L A	la i	 Lve	1													W C	re rec
	•				~, .	ړۍ	, AF	מץ ק	e Il	e C1	u I	le	Lys	Asp	G13	Ly	s Al	a G	lu V	ol Trp
			_ 13	17			12/	5			_									
	ATA	CTO	: c	AG (GGλ	GTG	GA	. GA	G AT	נו דיריית	3 Y m	200	~	144		-	15	3		162 C AGA
												~C	GAA	- AAA	CCA	GAC	; yc	A TO	TCC	C AGA
	Ile	Lev	G1	B (Gly	Val	Gli	2 G11	1 II	e Ph	0 T	VI	Glu	Tare	D					o Arg
	•		17									•		~, .	110	wsp	Th	r Se	r Pr	o Arg
	ATC	TIC	יו	γ <i>ι</i>	er i	CNC	180) >===		18	9			198			20	7		22.5
								AGC	FTC	3 AA	CY	λG	CIG	ATC	GAG	GCI	11	r ci	G AC	316 TAA 7
	Ile	Phe	Pb	e A	lla.	Gln	Ala	Arc			·									 z Asn
									,	· ns,	נים ני	y 3	val	Ile	Glu	Ala	Pho	Le	u Th	z Asn
			- 22	5			224													
		C1 G	GA	TA	CG	λλλ	λAG	XXX	GA.	CI	T	rc .	λAG	GIT	ACT	مملئ	261 C20			270 A GAG
	Pro	Val			·~~	7.100											GAL	- 66	A λλ	y Cyc
	_			,	***	rya	rys	LYS	Glu	i Lei	ı Pi	e.	ГУЗ	Val	Thr	Val	λsr	G1.	v Lv	s Glu
			27	Q			200													
	ATT (CCC	GIY	CT	CA,	λGλ	GTG	GAA	λλG	23	. Cr	T (~~~	306			315	;		324 3 AAC
	·			- -								·		ACG	GAC	ATA	GAC	GI	3 AC	S AAC
	ile)	PYO	Va.	l S	er .	λīģ	Val	Glu	Lys	λla	26	P I	Pro	Thr	ASD.	T10				
			333	1								_					ASD	Va.	Th	: Asn
7	מאכ כ	TG	λGZ	, \ λ'	TC (مكنت	342			351				360			369			378
-								TCT		TCC	CI	G }	W	GYY	GAA	GAC	CIC	AGZ	L AAJ	CAC
7	λx. Λ	/al	۸rg	, Il	le v	/al	Leu	Ser	Glu	Ser	ī.e.	 ., r	AFF 1							Asp
											-	٠.	ys .	GIU	GIA	Asp	Leu	yrg	Lys	λsp
G	ന്ദ്ര ദ	: A A	387		<u> </u>	:	396			405				414			477			
_				. Al		ATA (GAA.	CCT	TAC	YYY	CCC	GG	CA 2	AGA (GTC .	ATC	ATG	ATY	Gac	432 ATC
ν	al G	lu	Leu	11	e I	:le (:3,,													AIL
	al G				_			GIY	IYT	rÀ2	Pro	λ	la)	ra .	Val :	Ile	Met	Met	Glu	Ile
			441				I E N													
-	TG G	AC (GAC	TA	CT	AT 7	CAC	GAT	GGA	GAG	CTC	. G	GA C		~m.s		477			486
1.		:										_			·	TAT	TCT	CCY	GAG	λλG
_	eu A	י עב	чÞ	ТУ	TT	Ar 1	<u> 1</u>	yab	Gly	Glu	Leu	G	ג עו	la i	/al :		Ser	DT-	C1	
		4	195			5	DA.													
A	CG A	ra 1	TC	λG	λG	כ די אד	GG '	וכר י		513			S	22		!	531	•		540
-	CG A:									~	ICT	A.	NG T	cc c	TA J	VAG (GTG	CTT	CTC	TIC
T)	ur II	le F)he	λr	g V	al T	rp S	er :	Pro '	Val	Ser	T.s.								
												رب	, as 1,	¥ لاب	oi I	YS Y	Val	Len	1	Dha

Figure 14a_

		The) I'M	tog	_	arit	:im		-11 .					· • ·				
									471	TIED	450	(6	GP3) (cont	inue	đ)	
λ	AA 2	: אר כ	549	C11	.	558		•	567	7		57	76		E 0			
_					GAC .	ACA	Gλλ	CCG	TAC	CXC	c CI	T GI	NG AA	C AT	G GA	A TA	• AA	594 GGA
L	ys A	sn G	iy (Glu /	Asp :	Thr	61	D-C	~									GGA Gly
		_			_					GII	ı va	l Va	l As	n Me	t Gl	u Tyr	Lys	Gly
~-					·^^ (CG (3 77	GIT	GΥY	GCC	GA:	r cr	C GA	c GG	A GI	S G TTY	. TAC	648 CTC
A:	sn G	ly v	al o	מבן	lu A	la v	/al	Val	Glu	G1.	. ~		- :-					CTC
		£	57							U ₁	ردم	, ré	u As	p Gl	y Va	l Phe	Tyr	Leu
							-GA	AAG	ATC	yCy	ACA	AC(C GIY	CA:	ד ככי	TAT	TCG	702 AAA
TY	Œ G	ln L	eu G	lu A	sn T	Ar G	ly	Lys	Ile	λτσ	The							Lys
	٠	÷.	11		_			_				1111	· va	L AST	PEC	1AI	Ser	Lys
GC	ලේ	ות תי	AC G	CA A	. / AC A:	20 AC C		C) C	729			738	Β 3 Αλ1		747	,		756
									AGC	GCC	GIT	GIC	: AAT	, CII	, ecc	УСС	λCA	AAC
Al	a Va	נו וו	∕≖∘ λ	la A	BT A	sn G	ln (Glu	Ser	Ala	Val	Val	Asn					
		76	55		7.													
ٔ دع	A GA	A cc	ia To	33 GJ	ע א	rc c	AC J	AGG	783 GGA	~~		792	GAA		801			810
D							,				AAA	ATC	GAA	GGA	TAC	GAA	GAC	GCG
	, GI	n G1	УТ	JD GI	lu As	ш У л	Sp J	ra .	Gly	Pro	Lys	Ile	Glu	Glv	Tvr	Glu		
		M 3	Q.		~~													
ATA	L AT	C TA	TG	IK AI	א כא	C A	CA G	ecc (SAC.	ATC	ACA	846	CTC	~>>	855			864
Ile	Il	 1\sigma	 + G1	,										GAA	AAC	TCC	CCC	GTA
					e ul		е х	laj	(sp	Ile	Thr	CJÀ	Leu	Glu	Asn	Ser	Gly	Val
		87	3		0.0	_												
~~~	AAC	- AA	A GG	c cr	C TA	I CI	C C	GG (	י סדי	ACC (	GAA	GAA	λλC	ACG	909	CCX		918
Lys	Asr	Ly:	s G1	y Le		 r [_	 c	 ]										GGC
			_	-	4		<b>.</b> .	TY L	æu 1	inr (	Glu	Glu	Asn	Thr	Lys	Gly :	Pro (	Sly
		927	7		03	•												
			·	4 GG		TO	G C	vc c	TT G	TC (	EAA (	CIC	GGT	GIT	ACA	CAC (	י יידנד אינד	1/2 "እጥ
Gly	Val	Thr	Th	r Gly	/ Let	Se	r H	is 7.					Gly					
									•		oru 1	ceu	GIA	Val	Thr	Y eiH	/al H	lis
ATA	CTT	981 CCT		احلمك ر	990			9	99		10	800		1	017			36
							TA	C A	CA G	ec e	AC C	EAR I	CTC	GAT	AAA (	GAT 1	10 TC G	AG
Ile	Leu	Pro	Phe	: Phe	qaA :	Phe	Ty	T	nr G	 lv A		:1,.	 Leu ,	·				
		1035			1044													
AAG	TAC	TAC	λλC	TGG	1044 GGT	TAC	. G×	105 T A	53		_ 10	62		1	071		10	80
 									- T	AC C	TG T	TC /	ATG (	JIT (	.cc (	SAG G	GC A	GA
r\2	ТУI	LAr	λεη	Trp	Gly	Tyr	λв	p Pr	(I 0	/ L	 eu P	he M	Het \	 /-1 :				
					_						- •		(	ar I	rIO (	ilu G	ly A	rg

Figure 14b(Continued)

#### Thermotoga maritima Pullulanase (6GP3) (continued) 1089 1098 TAC TCA ACC GAT CCC AAA AAC CCA CAC ACG AGA ATC AGA GAA GTC AAA GAA ATG Tyr Ser Thr Asp Pro Lys Asn Pro His Thr Arg Ile Arg Glu Val Lys Glu Met 1143 GTC AAA GCC CTT CAC AAA CAC GGT ATA GGT GTG ATT ATG GAC ATG GTG TTC CCT 1152 Val Lys Ala Leu His Lys His Gly Ile Gly Val Ile Met Asp Met Val Phe Pro 1206 CAC ACC TAC GGT ATA GGC GAA CTC TCT GCG TTC GAT CAG ACG GTG CCG TAC TAC His Thr Tyr Gly Ile Gly Glu Leu Ser Ala Phe Asp Gln Thr Vai Pro Tyr Tyr 1260 TTC TAC AGA ATC GAC AAG ACA GGT GCC TAT TTG AAC GAA AGC GGA TGT GGT AAC Phe Tyr Arg Ile Asp Lys Thr Gly Ala Tyr Leu Asn Glu Ser Gly Cys Gly Asn 1314 GTC ATC GCA AGC GAA AGA CCC ATG ATG AGA AAA TTC ATA GTC GAT ACC GTC ACC Val Ile Ala Ser Glu Arg Pro Met Met Arg Lys Phe Ile Val Asp Thr Val Thr 1359 1368 TAC TGG GTA AAG GAG TAT CAC ATA GAC GGA TTC AGG TTC GAT CAG ATG GGT CTC --- --- --- --- --- --- --- --- --- --- ---Tyr Trp Val Lys Glu Tyr His Ile Asp Gly Phe Arg Phe Asp Gln Het Gly Leu 1413 1422 ATC GAC AAA AAG ACA ATG CTC GAA GTC GAA AGA GCT CTT CAT AAA ATC GAT CCA --- --- --- --- --- --- --- --- --- --- --- --- ---Ile Asp Lys Lys Thr Met Leu Glu Val Glu Arg Ala Leu His Lys Ile Asp Pro 1467 1476 ACT ATC ATT CTC TAC GGC GAA CCG TGG GGT GGA TGG GGA GCA CCG ATC AGG TIT --- --- --- --- --- --- --- --- --- --- --- --- ---Thr Ile Ile Leu Tyr Gly Glu Pro Trp Gly Gly Trp Gly Ala Pro Ile Arg Phe 1530 GGA AAG AGC GAT GTC GCC GGC ACA CAC GTG GCA GCT TTC AAC GAT GAG TTC AGA Gly Lys Ser Asp Val Ala Gly Thr His Val Ala Ala Phe Asn Asp Glu Phe Arg 1584 GAC GCA ATA AGG GGT TCC GTG TTC AAC CCG AGC GTC AAG GGA TTC GTC ATG GGA --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---Asp Ala Ile Arg Gly Ser Val Phe Asn Pro Ser Val Lys Gly Phe Val Het Gly

Figure 14C(Continued)

		Ther	mot.	oga	mar	d t in	MA .	Pull	ular		(6	GP3)	(c	ont	lzue	<b>4</b> )	
		162	9		167	D			-								
GG	A TA	c cc	A A	G GA	A AC	CAA	G AT	44 T	i.		165 ~~~	6 . ~ ~~		166	5		1674 TAC
												r Gr	r GG	A AG	CAT	AA A	TAC
G1	у ту	∓ Gl	у Гу	'3 G1	u Th	r Ly	s Il	e Ly	2 Yr	a CJ	y Va	l Va	l Gl	Se:	r Ile	 8 Ası	Tyr
		168	3	*	169	,		170									
GA	c cc	λ λλ	A CT	TK 7	C YY	A AG	r TT	C GC	- C CT	T-GA	ברר בי ידידי	N Baran		171!	, , ,		1728 TAC
														· AL	ATZ	· AAC	TAC
_ A3;	b GI	у Гу	s Le	u Il	e Ly	s Se	Ph	e XI.	a Le	u Ası	Pro	Gli	Glu	Th	116	) Asr	Tyr
		173	7		174	6		175	5		1764	1		1773	1		
تانات	A GC	G TG	r ca	C GA	C AX	CAC	: AC	y CIA	G TG	G GA	: 330	3 AAC	TAC	בייי:	, ecc		1782 AAA
Ala																	
																	Lys
CC	r (3)	1791 מגביי			1800	) 		1809	•		1818	1		1827			1836
				- AA	لمن د	TGG	AC	C GAJ	(CY)	(GAA	CIG	λλλ	AAC	GCC	CAG	λλλ	1836 CTG
Ala	ι λs _I	Lys	Ly	s Lvs	Gli	T	ጥኩ										Leu
		_									r reu	Lys	Asp	Ala	Gln	Lys	Leu
		1845	;		1854			1863	3		1872			1881			1000
	GG	r GCG	ATI	CI	CIC	ACT	JC1	CO	CCI	للملرئ ،	, CC	- Market	~~~	CAC	GGA	GGG	1890 CAG
Ala	Glv	/ Ala	714														
		/ Ala		. Let	. Den	inr	Ser	Gln	Gly	Val	Pro	Phe	Leu	His	Gly	Gly	Gln
		1899			1908			1017									
CYC	TTC	TGC	λœ	) ACG	ACG	AAT	TTC	λλC	GAC	AAC	שרר משנד	TAC	220	1935		:	1944
															CCT	ATC	TCG
Asp	Phe	Сув	λrg								Ser	Tyr	Asn	Ala	Pro	Ile	Ser
		1953			1962			1971		:	1980			1000		1	
ATA	AAC	GGC	TTC	GAT	TAC	CYY	AGA	AAA	CIT	CAG	TTC	ATA	GAC	GTG	TTC	AAT.	TAC
Ile	Asn	Gly	Pho														
		Gly						Lys	Leu	Gln	Phe	Ile	Asp	Val	Phe	Asn	Tyr
		2007			2016		:	2025		:	2034			2043		. 2	
CAC	AAG	GGT	CIC	λTλ	AAA	CIC	λGλ	AAA	GAA	CAC	CCT	GCT	TTC	AGG	CTG	AAA	770
His	Larg	Gly	1.00														
		Gly															
		2061		:	2070		. :	2079		2	880		•	2007			300
GCT	GAA	GAG	ATC	ХХХ	λλλ	CAC	CIG	Gλλ	TTT	CTC	CCG	GGC	GCG	AGA	AG A	ATTA	106
Ala	Glu	G3		 													
	-24	Glu	715	PAR	rys	nls	Leu	Glu	Phe	Leu	Pro	Cly	Gly	λrg	λrg	Ile	Val
	- 2	2115		2	2124		-	,,,,		_							
GCG	TTC	ATG	CIT	AAA	GλC	CAC	CCY_	GGT	CCT	GAT	CCC	TGG	AAA	C)C	) TO-	2	160
 11-	 bl -														71C	01G (	67G
~14	rne	Met	Leu	Lys	yab	His .	Ala	GJA	Gly	Asp	Pro	Trp	Lys	λερ	Ile	Val '	Val

Figure 14d(Continued)

	Thermotoga	maritima	Pullulanase	(6073)	(continued)	
ATT T	AC AAT CGA AJ	2178 NC TTA GAG	2187 AAG ACA ACA TAC	2196	2205 221 CCA GAA GGA AAA TG Pro Glu Gly Lys Tr	l 4 2G
AAT G	G GTT GTG AA	C AGC CAG 1	2241 NAA GCC GGA ACA	2250 GAA GTG	2259 226 ATA GAA ACC GTC GAA	8
GGA AC	A ATA GAA CTY	C GAT CCG C	2295 TT TCC GCG TAC	GTT CTG	2313	<b>L</b>

Figure 14e(Continued)

Figure 15a Thermotoga maritima MSB8 (Clone # 6GP2) Glycosidase

CTT TTA TTG ATC GTT GAG CTC TCT TTC GTT CTC TTT GCA AGT GAC GAG TTC Leu Leu Leu Ile Val Glu Leu Ser Phe Val Leu Phe Ala Ser Asp Glu Phe

GTG AAA GTG GAA AAC GGA AAA TTC GCT CTG AAC GGA AAA GAA TTC AGA TTC Val Lys Val Glu Asn Gly Lys Phe Ala Leu Asn Gly Lys Glu Phe Arg Phe

ATT GGA AGC AAC AAC TAC TAC ATG CAC TAC AAG AGC AAC GGA ATG ATA GAC Ile Gly Ser Asn Asn Tyr Tyr Met His Tyr Lys Ser Asn Gly Met Ile Asp

AGT GTT CTG GAG AGT GCC AGA GAC ATG GGT ATA AAG GTC CTC AGA ATC TGG Ser Val Leu Glu Ser Ala Arg Asp Met Gly Ile Lys Val Leu Arg Ile Trp

GGT TTC CTC GAC GGG GAG AGT TAC TGC AGA GAC AAG AAC ACC TAC ATG CAT Gly Phe Leu Asp Gly Glu Ser Tyr Cys Arg Asp Lys Asn Thr Tyr Met His

CCT GAG CCC GGT GTT TTC GGG GTG CCA GAA GGA ATA TCG AAC GCC CAG AGC Pro Glu Pro Gly Val Pne Gly Val Pro Glu Gly Ile Ser Asn Ala Gln Ser

GGT TTC GAA AGA CTC GAC TAC ACA GTT GCG AAA GCG AAA GAA CTC GGT ATA Gly Phe Glu Arg Leu Asp Tyr Thr Val Ala Lys Ala Lys Glu Leu Gly Ile

AAA CTT GTC ATT GTT GTG AAC AAC TGG GAC GAC TTC GGT GGA ATG AAC Lys Leu Val lle Val Leu Val Asn Asn Trp Asp Asp Phe Gly Gly Met Asn

CAG TAC GTG AGG TGG TTT GGA GGA ACC CAT CAC GAC GAT TTC TAC AGA GAT Gln Tyr Val Arg Trp Phe Gly Gly Thr His His Asp Asp Phe Tyr Arg Asp

GAG AAG ATC AAA GAA GAG TAC AAA AAG TAC GTC TCC TTT CTC GTA AAC CAT Glu Lys Ile Lys Glu Glu Tyr Lys Lys Tyr Val Ser Phe Leu Val Asn His

GTC AAT ACC TAC ACG GGA GTT CCT TAC AGG GAA GAG CCC ACC ATC ATG GCC Val Asn Thr Tyr Thr Gly Val Pro Tyr Arg Glu Glu Pro Thr Ile Met Ala

TGG GAG CTT GCA AAC GAA CCG CGC TGT GAG ACG GAC AAA TCG GGG AAC ACG Trp Glu Leu Ala Asn Glu Pro Arg Cys Glu Thr Asp Lys Ser Gly Asn Thr

CTC GTT GAG TGG GTG AAG GAG ATG AGC TCC TAC ATA AAG AGT CTG GAT CCC Leu Val Glu Trp Val Lys Glu Met Ser Ser Tyr Ile Lys Ser Leu Asp Pro

AAC CAC CTC GTG GCT GTG GGG GAC GAA GGA TTC TTC AGC AAC TAC GAA GGA Asn His Leu Val Ala Val Gly Asp Glu Gly Phe Phe Ser Asn Tyr Glu Gly

TTC AAA CCT TAC GGT GGA GAA GCC GAG TGG GCC TAC AAC GGC TGG TCC GGT Phe Lys Pro Tyr Gly Glu Ala Glu Trp Ala Tyr Asn Gly Trp Ser Gly

GTT GAC TGG AAG AAG CTC CTT TCG ATA GAG ACG GTG GAC TTC GGC ACG TTC Val Asp Trp Lys Lys Leu Leu Ser Ile Glu Thr Val Asp Phe Gly Thr Phe

CAC CTC TAT CCG TCC CAC TGG GGT GTC AGT CCA GAG AAC TAT GCC CAG TGG His Leu Tyr Pro Ser His Trp Gly Val Ser Pro Glu Asn Tyr Ala Gln Trp

GGA GCG AAG TGG ATA GAA GAC CAC ATA AAG ATC GCA AAA GAG ATC GGA AAA Gly Ala Lys Trp Ile Glu Asp His Ile Lys Ile Ala Lys Glu Ile Gly Lys

CCC GTT GTT CTG GAA GAA TAT GGA ATT CCA AAG AGT GCG CCA GTT AAC AGA Pro Val Val Leu Glu Glu Tyr Gly Ile Pro Lys Ser Ala Pro Val Asn Arg

ACG GCC ATC TAC AGA CTC TGG AAC GAT CTG GTC TAC GAT CTC GGT GGA GAT Thr Ala Ile Tyr Arg Leu Trp Asn Asp Leu Val Tyr Asp Leu Gly Gly Asp

GGA GCG ATG TTC TGG ATG CTC GCG GGA ATC GGG GAA GGT TCG GAC AGA GAC Gly Ala Met Phe Trp Met Leu Ala Gly Ile Gly Glu Gly Ser Asp Arg Asp

GAG AGA GGG TAC TAT CCG GAC TAC GAC GGT TTC AGA ATA GTG AAC GAC GAC Glu Arg Gly Tyr Tyr Pro Asp Tyr Asp Gly Phe Arg Ile Val Asn Asp Asp

AGT CCA GAA GCG GAA CTG ATA AGA GAA TAC GCG AAG CTG TTC AAC ACA GGT Ser Pro Glu Ala Glu Leu Ile Arg Glu Tyr Ala Lys Leu Phe Asn Thr Gly

GAA GAC ATA AGA GAA GAC ACC TGC TCT TTC ATC CTT CCA AAA GAC GGC ATG Glu Asp Ile Arg Glu Asp Thr Cys Ser Phe Ile Leu Pro Lys Asp Gly Met

GAG ATC AAA AAG ACC GTG GAA GTG AGG GCT GGT GTT TTC GAC TAC AGC AAC

Figure 15b (continued)

Glu Ile Lys Lys Thr Val Glu Val Arg Ala Gly Val Phe Asp Tyr Ser Asn

ACG TTT GAA AAG TTG TCT GTC AAA GTC GAA GAT CTG GTT TTT GAA AAT GAG Thr Phe Glu Lys Leu Ser Val Lys Val Glu Asp Leu Val Phe Glu Asn Glu

ATA GAG CAT CTC GGA TAC GGA ATT TAC GGC TTT GAT CTC GAC ACA ACC CGG Ile Glu His Leu Gly Tyr Gly Ile Tyr Gly Phe Asp Leu Asp Thr Thr Arg

ATC CCG GAT GGA GAA CAT GAA ATG TTC CTT GAA GGC CAC TTT CAG GGA AAA Ile Pro Asp Gly Glu His Glu Met Phe Leu Glu Gly His Phe Gln Gly Lys

ACG GTG AAA GAC TCT ATC AAA GCG AAA GTG GTG AAC GAA GCA CGG TAC GTG Thr Val Lys Asp Ser Ile Lys Ala Lys Val Val Asn Glu Ala Arg Tyr Val

CTC GCA GAG GAA GTT GAT TTT TCC TCT CCA GAA GAG GTG AAA AAC TGG TGG Leu Ala Glu Glu Val Asp Phe Ser Ser Pro Glu Glu Val Lys Asn Trp Trp

AAC AGC GGA ACC TGG CAG GCA GAG TTC GGG TCA CCT GAC ATT GAA TGG AAC Asn Ser Gly Thr Trp Gln Ala Glu Phe Gly Ser Pro Asp Ile Glu Trp Asn

GGT GAG GTG GGA AAT GGA GCA CTG CAG CTG AAC GTG AAA CTG CCC GGA AAG Gly Glu Val Gly Asn Gly Ala Leu Gln Leu Asn Val Lys Leu Pro Gly Lys

AGC GAC TGG GAA GAA GTG AGA GTA GCA AGG AAG TTC GAA AGA CTC TCA GAA Ser Asp Trp Glu Glu Val Arg Val Ala Arg Lys Phe Glu Arg Leu Ser Glu

TGT GAG ATC CTC GAG TAC GAC ATC TAC ATT CCA AAC GTC GAG GGA CTC AAG Cys Glu Ile Leu Glu Tyr Asp Ile Tyr Ile Pro Asn Val Glu Gly Leu Lys

GGA AGG TTG AGG CCG TAC GCG GTT CTG AAC CCC GGC TGG GTG AAG ATA GGC Gly Arg Leu Arg Pro Tyr Ala Val Leu Asn Pro Gly Trp Val Lys Ile Gly

CTC GAC ATG AAC AAC GCG AAC GTG GAA AGT GCG GAG ATC ATC ACT TTC GGC Leu Asp Met Asn Asn Ala Asn Val Glu Ser Ala Glu Ile Ile Thr Phe Gly

GGA AAA GAG TAC AGA AGA TTC CAT GTA AGA ATT GAG TTC GAC AGA ACA GCG Gly Lys Glu Tyr Arg Arg Phe His Val Arg Ile Glu Phe Asp Arg Thr Ala

Figure 15C(continued)

GGG GTG AAA GAA CTT CAC ATA GGA GTT GTC GGT GAT CAT CTG AGG TAC GAT Gly Val Lys Glu Leu His Ile Gly Val Val Gly Asp His Leu Arg Tyr Asp

GGA CCG ATT TTC ATC GAT AAT GTG AGA CTT TAT AAA AGA ACA GGA GGT ATG Gly Pro Ile Phe Ile Asp Asn Val Arg Leu Tyr Lys Arg Thr Gly Gly Met

TGA 1991

END

Figure 15d(continued)

# Figure No. 16 Thermotoga maritima MSB8(6gb4)

_	. •••		- A	ra II	i C G	C CI	G AA	I GG	TT(	C TG	G AG	C GT	T AG	GA	44 T	C GN	A CC	~				
1	. Me	t L	'S A:	rg II	le As	p Le	u As:	n Gl	y Phe	e Tr	Se	r Va	1 200	. D.	-	- 63		G AC	A T	T	TCG	•
												- ·	- /;	, na	h wa	n GI	u GI	y Ar	g Pi	1e	Ser	2
61	لمل	T C																				
	- 11	I GA	LA GO	GG AC	T GI	rg cc	A GG	GT.	CTC	CAC	3 CC1	A GA	ר כדנ	GT	C AG	A AA	A GG	тст	т С <del>л</del>	<b>-</b>	CCN	
21	Ph	e Gl	u G	ly Th	r Va	l Pr	o G1	y Val	[ Va]	Glr	Ala	ASI	Let	ı Va'	1 A	7 130					- CCA	12
												•			• ~~;	y Ly.	2 GI	λ re	uLe	u	Pro	4
121	CA	ר רר	ים די		~ ~																	
41	u.	0 D-		AC GI	1 60	G AT	G AA(	GA	A GAT	CTC	TTC	AAC	GAA	AT	A GA	A GAG	C AG	A GA	G TG	G	ATC	18
4.	n1	S PI	0 13	r Va	1 G1	у Ме	t Ası	Glu	Asp	Lev	Phe	Lys	Glu	Ile	Gl	ı Ası	o Ar	g (G)	1 Tr	_	71.0	
																•		, ••		. ب	116	6
181	TA	C GA	G AG	G GA	G TT	C GA	ם דדר	• •	CAN							•						
61	Tv	r Gl	13 Ar	G GA	Db			- 222	GAA	GAT	GTG	AAA	GAG	GGG	GA	CG1	CTC	GA'	CI	C	GTT	24
	-			g Gl	u Pii	e GI	u Pne	Lys	Glu	Asp	Val	Lys	Glu	Gly	Glu	a Arg	Va:	L As	Le Le	u 1	Val .	8
				_																		_
241	TT	I GA	G GG	C GT	C GA	C AC	CTC	TCG	GAT	GTT	TAT	<u>Ст</u> с										
81	Ph	e Gl	u Gl	y Va	l As	o Th:	C Len	Ser	Acn	Va.	T		~~~	661	G1-1	TAC	CT	r GG	A AG	ני	/CC	30
				y Va				261	vsh	vai	ıyr	ren	Asn	Gly	Val	Тух	Let	Gl	/ Se	r 1	ŕhr	10
201											•											
301	GA	A GA	C AT	G TT	C AT	C GA	TAT	CGC	TTC	GAT	GTC	ACG	AAC	GTG	TTG		G D I					2.5
101	Glı	ı Asl	p Me	t Ph	e Il	e Glu	Tyr	Arg	Phe	Asp	Val	Thr	Asn	Val	7.00	7	C1.		, AA.		-AC	35
										•				•41	Den	Lys	GIU	Lys	ASI	7 F	iis	120
361	CTC	. 220	:	C 700	~																	
121	Tax			G TA	C AT	A AAA	TCT	CCC	ATC	AGA	GTT	CCG	AAA	ACT	CTC	GAG	CAG	AAC	TAC	: G	:GG	420
	Det	. ny:	s va	l Ty:	r Ile	e Lys	Ser	Pro	Ile	Arg	Val	Pro	Lys	Thr	Leu	Glu	Gln	Asn	TVI		יוני	140
															•			•		_	-3	
421	GTC	CTO	GG	C GG:	ר ככי	GAA 1	GAT	CCC	ልጥሮ	a Ca	CCN	m> c										
141	Val	Let	. G1	C GG:	, Dec			200	71-	AUA	GGA	TAC	ATA	AGA	AAA	GCC	CAG	TAT	TCG	T	AC	480
				y Gl	, ,,,	, GIO	ASD	PIO	TTE	Arg	Gly	Tyr	Ile	Arg	Lys	Ala	Gln	Tyr	Ser	T	уr	160
481	GGA	TGG	GA	TGG	GG1	GCC	AGA	ATC	GTT	ACA	AGC	GGT	АТТ	TCC	מממ		~~~			_		
161	Gly	Trp	Ası	Trp	Gly	Ala	Arq	Ile	Val	Thr	Sar	Gly	73.0	T		-	GIC	TAC	CTC	G	AG	540
					•							GLY	116	irp	ьуs	Pro	Val	Tyr	Leu	G	lu	180
541	CTC																					
	616	TAC	AGC	GCA	CGI	CII	CAG	GAT	TCA	ACG	GCT	TAT	CTG	TTG	GAA	CTT	GAG	GGG	444	G	ΔT	600
181	Val	Tyr	Arg	Ala	Arg	Leu	Gln	Asp	Ser	Thr	Ala	Tyr	Leu	Leu	Glu	T.ess	G) u	C1	7	3.		
												-				200	.014	GLY	Lys	A	sp	200
601	GCC	CTT	GTG	. ארכ																		
201	בוג	1		AGG	. 616	AAC	GGT	TTC	GTA	CAC	GGG	GAA	GGA	TAA	CTC	ATT	GTG	GAA	GTT	T	AT	660
	~~~	neri	vai	Arg	Val	Asn	Gly	Phe	Val	His	Gly	Glu	Gly	Asn	Leu	Ile	Val	Glu	Val	T	vr	220
																					· -	
661	GTA	AAC	GGT	GAA	AAG	ATA	999	GAG	ىدىلىك	درس.		~~	~									
221	Val	Asn	Glv	GAA	1	T1.	21	0 70		-	GTT.	CTT	GAA	AAG	AAC	GGA	GAA	AAG	CTC	T	C	720
			- - 3	Glu	Lys	116	GIÀ	GIU	ru6	Pro	Val	Leu	Glu	Lys	Asn	Gly	Glu	Lys	Leu	Pì	1e	240
20.								•														
721	GAT	GGA	GTG	TTC	CAC	CTG	AAA	GAT	GTG .	AAA	CTA-	TGG	ጥልጥ	CCC	TCC							
241	Asp	Gly	Val	Phe	His	Léu	Lvs	Asn '	Val	live					- 66	AAC	GTG	GGG	AAA	CC	:G	780
							-,-	٧,	- 44	~y5 .	u=u	TIP	ıyr	PTO	Trp	Asn	Val	Gly	Lys	Pr	:0	260

	781	TAC	CIG	TAC	GAT	TTC	سدن															•
	261	Tyr	Leu	Tur	A	246	GII	1-1-0	GTC	TIC	S AA	A GAC	TTA	AAC	GGA	GAG	ATC	TAC			A GAA	
		-		-3-	Asp	Pne	Val	Phe	Va]	Let	ı Lys	Asp	Leu	Asn	Glv	Glu	Tle	75.00	. AGA	GAJ	A GAA	840
	.																116	Tyt	Arg	Gli	Glu	280
	841	AAG	AAA	ATC	GGT	TTG	AGA	AGA	GTC	מאל י	אדר											
:	281	Lys	Lys	Ile	Gly	Leu	Aro	Aro	Val	han		GIT	CAG	GAG	CCC	GAT	GAA	GAA	GGA	AAA	ACT Thr	900
					•		9	Arg	val	Arg	Ile	Val	Gln	Glu	Pro	Asp	Glu	Glu	Glv	Tare	Th-	
	901																					300
		-110	ATA	TTC	GAA	ATC	AAC	GGT	GAG	AAA	GTC	TTC	CCT	220	GGT							
- -	101	Pue	Ile	Phe	Glu	Ile	Asn	Gly	Glu	Lvs	Val	Pho	33-	-	GGŢ	GCT	AAC	TGG	ATT	CCC	TCA	. 960
				•						-20		r ne	AIZ	Lys	GGT Gly	Ala	Asn	Trp	Ile	Pro	Ser	320
و ٠	61	GAA	AAC	ATC	~~~	١.٥٥																
. 3	21	G3 11	Acn	71.	-	ACG	TGG	TTG	AAG	GAG	GAA	GAT	TAC	GAA	AAG	CTC	CTC					
			nsii	TIE	ren	Thr	Trp	Leu	Lys	Glu	Glu	Asp	Tyr	Glu	AAG Lys	Tau	U- 1	~~A	ATG	GCA	AGG	1020
													•		-, -	Deu	val .	гλг	Met	Ala	Arg	340
10	21	AGT	GCC .	AAT .	ATG .	AAC	ATG (CTC	NGC.						ATC '							
3	41	Ser .	Ala.	Asn I	Met :	Asn A	det :		NGG	GIC	TGG	GGA	GGA .	GGA .	ATC '	TAC (GAG 2	AGA	GAG	ATC	TTC	13000
								Leu .	Arg	Val	Trp	Gly	Gly (Gly	ATC 1	Tyr (3lu j	Ara (G7 1,	71.	nh-	1080
108	27 .	TD - 1																		-16	FIIE	360
36		1AC /	AGA (CTC 1	rgr (SAT C	AA C	TC (GT :	ATC .	ATG (GTG '	ree e	~n~ ~	GAT 7							
36	,1	ryr 1	rd I	Leu (ys A	ksp G	lu I	eu (ly :	Ile :	Met 1	Val 1		-AG (AT T	TTC 3	TG I	CAC	GCG :	rgt -	CTT	1140
									-				LIP C	etu 1	SAT 7	he M	et I	yr 1	lla (Dys :	Leu	380
114																						
38	1 0	lu T	vr p	a	T	-71 C	II C	CG T	GG 7	TTC)	AGA A	AAA C	TC G	CG A	AC G	AA G	AG G	מ בר	מ מ			
				7	n de.	15 L	eu p	ro T	tb E	he A	Arg I	ys L	eu A	la A	AC G	lu G	מ ווי	ו או		WG 7	AIT	1200
120																						400
120	⊥ (։ 	TG A	GA A	AA C	TC A	GA T	AC C	AT C	CC T	CC A	TT G	TT C	T C T	aa -	GC G							
40	ı v	al A	rg L	ys L	eu A	rg Ty	YT H	is P	ro s	er T	ים פו	'al t		GG 1	GC G	SA A	4C Y	-C G	A AA	AC A	AC	1260
									_		- ·	at D	eu T	בס כב	ys G	ly A	an As	sn G	lu A	sn A	sn	420
126	L T	GG G(SA T	בר פֿי	T (1			_														
423	T	rp G	v Pi	D		- TC	G GC	SA AJ	AT A	TG G	CC A	GA AJ	AA G	rg gj	AT GO	T A	C Az	ر	TC C			
		•		TE AS	p G	iu Tr	7P G1	y As	א מפ	et A	la A	rg Ly	/S Va	al As	AT GO	רד ע	- h			3A A.	AC.	1320
																.,	CAS	orr Tre	eu G.	Ly A	SD	440 .
1321	AC	G CI	C TA	C CI	C TI	'C GA	T TI	T CC	T G	AG AT	יר די	~m ~c		<u>.</u> .	LA GA							
441	Ar	g Le	u Ty	T Le	u Ph	e As	p Ph	e Pr	. G			31 66	C GA	LA GA	LA GA	c cc	G TC	C AC	T CC	C T	AT.	1380
							•		0 01	. 4 11	re Cy	/S Al	a Gl	n CJ	LA GA Lu As	p Pr	o Se	r Th	r Pr	O T3	/r	460
1381																						
461	Tr	יים מ		- 2		A TA	C GG	C GG	T GA	AA A	A GC	C AA	C AG	C GA	A AA	G CZ	A CC			_		•
			J 3E	ı se	r Pr	o Ty:	r Gl	y Gl	y Gl	u Ly	s Al	a As	n Se	r 63	A AA u Ly			A GA	C AG	G CA	C :	1440
															- <i>-</i>	9 GT	n GT	y As	p Ar	g Hi	s	480
1441	GT	C TG	TA	CGT	G TG	G AG1	GGr	TC/	2 N.T.	~ ··		_			GAJ							
481	Va:	l Trj	Ту	r Va;	l Tro	Ser	. 61.		3 M1	G AA	C TA	C GA	A AAC	TA	C GAJ	A AA	GAC	ACC	C GG	A AG	G 1	.500
							. 61)	TI	Me	t As	n Ty	r Gl	ı Ası	туз	GAJ Glu	Lys	Asr	Thi	r G):		~	
1501	TT	, b.m-														-	r			, ,,,,	3	500
501	Dh -	- WI.C	AGC	GAC	TI	. GGA	TII	CAC	GG	r GC:	r ccc	CAT			ACG							
-01	≥n€	: Ile	Ser	Glu	Phe	Gly	Phe	Gln	Glv	/ Al=	a Pro	י אור		GAG	ACG	ATA	GAG	TTC	LIT	TC	A 1	560
						•			•			- 4115	PTO	Glu	ACG Thr	Ile	Glu	Phe	Phe	Sez	•	520
1561	AAA	ccc	GAG	GAA	AGR	C		_														
521	Lys	Pro	Gli	63	N	GAG	ATA	TTC	CAT	CCC	GTC	ATG	CTG	AAG	CAC	AAC	444	C>-		~		
				914	AIG	Glu	Ile	Phe	His	Pro	Val	Met	Leu	Lys	CAC His	Aen	L	CAG	GTG	GAA	. 1	620
								1	Fign	re :	16h/		inue	. , ,	3	W91)	rys	Gln	Val	Glu	1	540
								_			(COUL	TURE	2 a)								

1621	GG	A CA	G GA	A AG	א דד	C 5.00	~														
541	. G1·	י מו	n G)			G AT	CAG	3 TTC	AT	TTC	GGZ	AA:	I II.	T GG	A AA	G TG	T AA	A GA	т тт	C GAC	
		, 01.	01	u AI	g Le	u Ile	e Arc	Phe	: Ile	Phe	Gly	' Ası	n Phe	e Gl	Lv:	s Cv	9 T.v.	c ha	- n.	C GAC e Asp	
																					56
1681	AG:	r TT	r GŢ	G TA	CT	GTC	CAC	CTC	. אאכ	. Cac				_							
561	Sei	Phe	Va:	l Ty:	r Let	ı Ser	Gl n	T.e.	Aco	- C1-	33-	GAU	GCC	3 ATC	: AA	TT	GG:	CT	r gaj	A CAC	1740
			•						- ASII	Gin	ATA	Glu	ı Ala	lle	Ly:	Phe	e Gly	/ Va:	Gli	A CAC	580
1741																					
581	Trr	220	Car	- AGC	AAC	TAC	: AAA	ACG	GCC	GGC	GCT	CTC	TTC	TGG	CAG	TTC	. AAC	GAC		TGG	
		, ALS) Sei	Arg	Lys	Tyr	Lys	Thr	Ala	Gly	Ala	Leu	Phe	Trp	Gln	Phe	Acn	2000		Trp	1800
																					600
1801 601	CCG	GTC	TTC	AGC	TGG	TCC	GCA	GTC	GAT	TAC	TTC		100						•	*	
601	Pro	Val	Phe	Ser	Trp	Ser	Ala	Val	Asn	There	Dha	7	AGG	CCC	AAA	GCT	CTC	TAC	TAC	TAT	1860
601									···p	- 7 -	FILE	ьys	Arg	Pro	Lys	Ala	Leu	Tyr	Tyr	Tyr	620
1861																					
621	Ala	Aro	Arc	Dha	770	GCT	GÁA Glu	GTT	CTA	CCC	GTT	TTG	AAG	AAG	AGA	GAC	AAC	AAA	ATA	a a d	1920
		5	n g	Pne	Pne	Ala	Glu	Val	Leu	Pro	Val	Leu	Lys	Lys	Arg	σεA	Asn	Lvs	Tla	61	
																					640
1921	CTG	CTG	GTG	GGT	GAG	CGA	TCT Ser	GAG	GGA	GAC	444	AC A	h.cm	-							
641	Leu	Leu	Val	Gly	Glu	Arg	Ser	Glu	Glv	Aen '	Laro	7	AGI	CTC	TCT	CAG	GCT	TGC	AGC	CTA	1980
						•			,	ASP.	bys .	Arg	ser	Leu	Ser	Gln	Ala	Cys	Ser	Leu	660
981	CGA	GAA	GAA	GGG	202																
661	CGA Arg	Glu	Gl.	Class	AGA	AAA	GGT .	ATT	CGA .	AAA (SAC '	TTA	CAG	AAC	GGT	ACT	CCC	AGC	AGA	CGG	2040
	Arg			Gry	Arg	rys	Gly	Ile 2	Arg 1	Lys)	l qa	Leu	Gln	Asn	Gly	Thr	Pro	Ser	Ara	Ara	680
																			- 7	9	880
041	TGT					20	55														٠.
681	Cys	Glu.	Phe.	Gly	End	68	5														

Figure 16C(continued)

Figure No. 17 Bankia gouldi (37gp4)

	1 A'	TC .																								
	- A	IG A	AA A	AA J	LAT	CTA	CT	AT	G TI	TA	AA	AGG	CIT	ACC	G TA	T C	TA	CCT	יאריים		TT.					
	1 M	et L	ys L	ys A	\sn	Leu	Lev	Me	t Ph	e L	ys	Arq	Leu	Th	r 150	- I		D			TT)	TA	G	TTG	60)
											•				y	- 1	eu	PTO	Leu	Phe	Let	1 Me	t I	Leu	20	,
6:	ı c	C T	רם מר	· ሞእ *	~~																					
21	 I Ta			IA A	rc.T.	TCA	GTA	GC	CA	A T	CI	CCT	GTA	GAZ	AA A	A C	AT	GGC	CGT	TTA	CAR	, C.T.				
2.		2U 50	er L	eu S	er	Ser	Val	Ala	Gl	n S	er	Pro	Val	Glu	ı Lv	s H	is	G) v	Arc	Tax	CAR Gln			AC	120	
															•			,	AL 9	reu	GIR	va.	LA	sp	40	
121	GC G1	A A	C C	GC A	TT	CTT	AAT	GCC	· •	T ~	~ n	~							•							
41	. G1	y As	n A	ra I	l e	Leu	Aen	21-		- 01	374. 1		ATT	ACG	AG	C T	ra (GCT	GGT	AAC	AGC	CTC	Ť	TT	180	
•							nsu	M.	. se	r G	Ly (Glu	Ile	Thr	Se	r Le	eu j	Ala	Gly	Asn	AGC Ser	Leu	P	he	60	
101																										
181	. 16	G AC	IA TE	AT G	CT .	GGA:	GAC	ACC	TC	G G#	T 3	TTT	TAT	AAT	GC.	A GA	A.ZA. J	ر ب	CTT	C 3 T	TTT					
61	Tr	p Se	T As	an A	la	Gly	Asp	Thr	Sea	. As	p I	Phe	Tyr	Asn	Δ1:			~~		GAT	TTT Phe	TTA	G	CA	240	
											_		•		-		ш.	lnr	vai	Asp	Phe	Leu	A.	la	80	
241	GA	A AA	C TO	A D	ב דמ	ACC.	~~~	~~~																		
81	G1	u As	ייי די	~ N		^	- CA	CIT	ATT	AG	A, A	ATA	GCT	ATG	GG	GT	'A 2	LAA	GAA	AAT	TGG	GAT	G	GC	300	
		- /10	** 11	.р ж	sn :	ser	Ser	Leu	Ile	: Ar	g I	le	Ala	Met	Gly	, Va	1 1	ys	Glu	Asn	TGG Trp	Asp	G1	112	100	
																									100	
301	GG.	A AA	T GO	C T	AT A	ATT	GAT	AGT	CCG	CA	GG	AG	Caa	445	CCT						GTT					
101	Gl	y As	n Gl	y Ty	yr 1	Ile	Asp	Ser	Pro	GI	n G	3	G1 5	23	33-	-	Α А	att .	AGA	AAA	GTT Val	ATT	G.	T	360	
													GIII	GIU	ATA	Ly	s I	le.	Arg	Lys	Val	Ile	As	Ď.	120	
361	GC	٠	T 2 T		· ·																	•	•			
121	۵۱:			1 60	JT F	AAC	GGC	ATA	TAT	GT.	A A	TA .	ATA	GAC	TGG	CA	C A	CT (CAC	GAA	GCA	GAG	7-7	'a	420	
	NI.	. Al	a 11	e Al	a A	lsn (Gly	Ile	Tyr	Va	1 1	le .	lle	Asp	Trp	Hi	s T	hr 1	His -	Glu	GCA Ala	61	T.0			
																									140	
421	. TAC	: AC	A GA	T GA	G G	CT (STT	GAC	TTT	77-	י ב	CC :	- C D	h TC	~~~		_								• .	
141	Tyr	Thi	As	p Gl	uА	la 1	/al	Asp	Phe	Dha	771	h- 1		MIG.	GCA	GAC	C C	TA :	rac (GGA	GAT Asp	ACT	CC	C .	480	
							_			****	- +,	111 /	ug i	met	Ala	Asp) L	eu 2	(At (Gly .	Asp	Thr	Pr	0	160	
481	דממ	دشت				_																				•
161	7	77-7		J TA	TG	AA A	TT	TAŢ	AAC	GAC	C	T A	TA :	TAC	CAA	AGI	T	G C	CT (TT.	ATT .	DAA	• ממ	r	540	
	Vall	val	. Met	Ту	r G	lu]	le	Tyr	Asn	Glu	P	to 1	le :	Tyr	Gln	Ser	T	GD \$	ro T	/all	ATT :		A	_		
																		-			**E .	Lys	A 5 1	13.	180	
541	TAT	GCA	GAC	CA.	A G	TA P	TT (GCT	GCT	מדמ	CC	ייי ייי	·~~ .								ATA I			•		
181	Tyr	Ala	Gli	1 Gl:	n V	al I	le :	Ala .	63	71.			-1 -	AAA I	GAC	CCA	G.	A TA	AT 7	TA 2	ATA I	ATT	GT	A	600	
									Gly	116	AI	gs	er I	Lys .	Asp	Pro) As	Sp A	sn I	eu :	ATA :	Ile '	Va]	L	200	
601																										
201	001	AC1	AGC	: AA:	T T	AT T	CT (CAG	CAA	GTT	GA	TG	TA C	CA :	TCA	GCA	. GA	יכ כ	מ מי	י מידי	rc r (- N. M	. ~			
-01	Gry	Inr	Ser	Ası	T)	yr S	er (3ln (Gln	Val	As	p۷	al A	la s	Ser	Ala	Ac	n D			CT (Ser A	3VI 1			660	
												•						, P	20 1	TE S	er,	rsb .	rnr	•	220	
661	AAT	GTG	GCA	TAT	C AC	יי די	TA C	י דמי	متماسط	T > ~																
221	AAT Asn	Val	Ala	Tvi	ተጉ	יים דיים	D11 "		· • •	w. TWI	GC.	A G	CA 1	TT ;	4AC	CCG	CA	T G	A TA	AC 1	TA A	GA J	AAT	•	720	
	Asn			- , •	- ••		cu r	145 1	rne	ıyr	ΑŢ	аА	la P	he A	Asn	Pro	Hi	s A	sp A	sn I	eu A	rg)	Asn		240	
	GTA Val	GCA	CAG	ACA	GC	AT	TA G	AT A	LAT :	AAT	GT'	T G	тт	TG 7	ידיידיי	سدت	n.c	» c								
241	Val	Ala	Gln	Thr	: Al	a L	eu A	sp A	sn i	naA	Va:	l a'	la t	۰. ۰	ha '	U. 3		A G	na T	GG G	GT A	CA A	TT		780	
								-						cu F	ne	val	Th	r G	lu T	rn G	י או	h- 1	۰,		260	

71	'81 TTA AAT ACC CON CON	
	THE ACC GGA CAA GGA GAA CCA GAA	
	61 Leu Asn Thr Gly Gln Gly Glu Pro Asp Lys Glu Ser Thr Asn Thr Trp Met Ala Phe Leu	840
	Ash Thr Trp Met Ala Phe Leu	280
84	41 AAA GAA AAA GGT ATA AGT CAG GGT	•
28	41 AAA GAA AAA GGT ATA AGT CAC GCT AAT TGG TCT TTG AGT GAC AAA GCT TTT CCT GAA ACA	•
	81 Lys Glu Lys Gly Ile Ser His Ala Asn Trp Ser Leu Ser Asp Lys Ala Phe Pro Glu Thr	900
		300
90	Ol GGG TCT GTA GTT CAA GCA GGA CAA GGT GTA TCT GGT TTA ATT AGC AAT AAA CTT ACA GCC	
30	Gly Ser Val Val Gln Ala Gly Gln Gly Val Ser Gly T	960
	Set Gly Leu Ile Ser Asn Lvs Leu The	320
96:		
321	TCT GGT GAA ATT GTA AAA AAC ATC ATC CAA AAC TGG GAT ACA GAG ACC TCT ACA GGA CCT 1	
•	1 Ser Gly Glu Ile Val Lys Asn Ile Ile Gln Asn Trp Asp Thr Glu Thr Ser Thr Gly Pro	020
• • • •	•	340
1021	ACA CAA TGT AGT ACT ATA GAA TCT ATT	
341	1 Lys Thr Thr Gln Cys Ser Thr Ile Glu Cys Tla No. 210 ATG GAA ACA GCA CAA GCA 10	080
	and all Ala Met Glu Thr ala Cla	360
1081		
361	GGA GAT GAA ATT ATA ATT GCC CCT GGA AAC TAC AAT TTT CAA GAC AAG ATA CAA GGT GCC 11	
	ASH TYP ASH PRE GIR ASP ING THE CITY OF	40
		80
1141	THE COL AGI GIT TAC CTT TAT GGT ACT COT	
381	Phe Asn Arg Ser Val Tyr Leu Tyr Gly Ser Ala Asn Gly Asn Ser Thr Asn Pro Ile Ile 4	00
	4 deposit And Ash Gly Ash Ser Thr Ash Pro Ile Ile	00.
1201	·	
401	TTA AGA GGC GAA AGC GCT ACA AAC CCT CCT GTT TTC TCA GGA TTA GAT TAT AAC AAT GGC 120	60
	val Pne Ser Gly Leu hen Trees have	
1261		20
421	TAC CTA TTA AGT ATT GAA GGT GAT TAT TGG AAT ATT AAA GAT ATA GAG TTT AAA ACT GGG 132	
421	Tyr Leu Leu Ser Ile Glu Gly Asp Tyr Trp Asn Ile Lys Asp Ile Glu Phe Lys Thr Gly	20
	44	0
1321	TCT AAA GGT ATT GTT CTT GAC AAT TCT	
441	TCT AAA GGT ATT GTT CTT GAC AAT TCT AAT GGT AGT AAA TTA AAA AAC CTT GTT GTT CAT 138	0
	Ser Lys Gly Ile Val Leu Asp Asn Ser Asn Gly Ser Lys Leu Lys Asn Leu Val Val His 46	
1381		•
461	GAT ATT GGA GAA GAA GCT ATT CAC TTG CGT GAT GGA TCT AGC AAT AAT AGT ATA GAT GGT 144	
	Asp Ile Gly Glu Glu Ala Ile His Leu Arg Asp Gly Ser Ser Asn Asn Ser Ile Asp Gly 48	0
		0
1441	TGC ACT ATA TAC AAT ACA GGT AGA ACT AAA CCT GGT TTT GGT GAA GGT TTA TAT GTA GGC Cys Thr lle Tyr Asn Thr Gly Arg Thr Lys Pro Cly Pro Cys Thr Lys Pro Cys Thr L	
481	Cys Thr Ile Tyr Asn Thr Gly Are The Live Tyr GGT GAA GGT TTA TAT GTA GGC 1500	
	Cys Thr Ile Tyr Asn Thr Gly Arg Thr Lys Pro Gly Phe Gly Glu Gly Leu Tyr Val Gly 500	
501	TCA GAT AAA GGA CAA CAT GAC ACT TAT GAA AGA GCT TGT AAC AAT AAC ACT ATT GAA AAC 1560	
•	THE CYS ASD ASD ASD ASD TO THE TAX	
)
1561 7	TGT ACC GTT GGA CCC AAT GTA ACA GCA GAA GGC GTA GAT GTT AAG GAA GGT ACA ATG AAC 1620	
521 (Cys Thr Val Gly Pro Asn Val Thr Ala Glu Gly Val Asp Val Lys Glu Gly Thr Met Asn 540	
	540 Thir Ala Glu Gly Val Asp Val Lys Glu Gly Thr Met Asn	

Figure 17b(continued)

	162	21 A	CI	ATT	r At	A A	GA A	ልጉ ጉ	GC O																		
	54	1 T	hr	Ile	: I1	e A	ra A	SD C	ve v	1G 1	TT C	TCT	ပေ	A GAJ	A GG	A A	II I	CA	GGA	GAA	. AA	TA	GC	TCA	GAT	168	^
		l T				•	•		,	aı F	ne .	Ser	Ala	Glı	1 G1	y I	le s	er	Gly	Glu	As	n s	er .	Ser	Ast	56	
	168																										,
		1 A	la	Phe	T1.		1 T	ra a	AA G	GA G	CC :	TAT	GGT	TI	GT	A TA	C A	GA 2	AAC	ACG	TT	T A	1T (مصدت	GAT		
		1 A	-	• 116	11	e As	ip Le	eu Ly	YS G	ly A	la :	lyr	Gly	Phe	Va.	1 Ty	T A	rg)	Asn.	Thr	Phe		en T	311 751	GAT	1740	
																										580)
	174	ı G	3T	TCT -	GA	A GI	'A A'	TA AI e As	AT A	T G	GA C	STA	GAC	TIT	TT	A GA	TAC	3A 6	TT-	202			_				
	58:	ı G	У	Ser	Glı	i Va	1 11	e As	n Tì	r G	ly v	/al	Asp	Phe	Let	ı As	D A	ra 6	11.	カレル	GGA	TI	T A	LAT	ACA	1800	
																										600	
	180	r GC	T	TTT	AGZ	AA A	T GC	A AT a Il	A TI	T G	A A	AT .	ACA	TAT	AAC	-بات	T CC			:							
	601	L G1	у	Phe	Arg	, As	n Al	a Il	e Ph	e G1	u A	sn'	Thr	Tvr	Asn	· Ci		· A	GT I	AGA	GCT	TC	A G	AA	ATT	1860	
																	4 61	.у 5	er /	urg	Ala	Se	∓ G	lu	Ile	620	
	1861		A A	CI	GCI	CG.	T AA	A AA S Lv	A CA	A GG	тт	/ است	CCT	CNN	~~~												
	621	Se	r 1	hr	Ala	Arg	J Ly	s Ly	s Gl	n Gl	v S	er 1	Pro	GAA	CAA	ACT	r CA	C G	II 1	CGG	GAT	AA'	I A	TT .	AGA	1920	•
														GIU	GIN	Tni	Hi	s V	al 1	LÞ	Asp	Ası	n I	le.	Arg	640	
	1921	AA	c c	CI	AAT	TC	GT:	T GA	ידי ינ	T CC	h h-	.															
	641	Ası	n P	ro	Asn	Sei	. Va	T GA:	Ph	e Pr	A A.	LA F	AGT	GAT	GGT	ACA	GA.	A A	AT C	TA	GTA _.	AA:	. Az	AA :	TTC	1980	
		Ası						•	55		- 1.		er.	Asp	Gly	Thr	Gl	u As	sn L	eu '	Val	Ası	L	/s 1	Phe	660	
	1981	TG	- 0	CA I	GAT	TGG	"אא	ידע י																			
	661	Cys	P	ro .	Asp	Trp	Asr	T ATA	· Gl		A TO	T A	LAT	CCT -	GTA	GAC	GAJ	A AC	C A	AC (CAA	GCA	CC	T	ACA	2040	
						•		lle	. 01(. F.L	J Cy	SA	sn	Pro	Val	Asp	Gli	1 Th	r A	sn (Gln	Ala	Pr	o 1	hr	680	
;	2041																										
	681	Ile	Se	er 1	Phe	Len	Sar	CCT	GIT	· AAC	: AA	TA	TT A	ACT :	TTA	GTT	GAA	GG	T T	AT A	LAT	TTA	CA	A G	TT	2100	
							261	Pro	VET	AST	As	n I	le 7	Chr 1	Leu	Val	Glu	Gl	у Т	Yr A	rsn	Leu	G1:	n V	al	700	
:	2101																										
	701	GAA Glu	Va	1 2	en.	NI.	ACT	GAT	GCA	GAT	GG.	A A	CT A	TT C	EAT	AAT	GTA	. AA	A CT	T T	AT :	ATA	GA'	т в	a C	2160	
		Glu				~1a	Inr	qaA	Ala	Asp	G1	y Tì	hr I	le A	Asp .	Asn	Val	Ly	s Le	u T	yr :	Ile	Ası	 - A:	sn.	720	
	721	AAT Asn	T.o	A G	TT.	AGG	CAA	ATA	AAT	TCT	AC.	T TC	CA T	AT A	AA '	TGG	GGC	CA:	rTo	TG	T 7	г с-т -	CC		N -TP	2220	
		naA		u v	aı.	Arg	GID	Ile	Asn	Ser	Thi	Se	er T	yr L	ys :	Trp	Gly	His	s Se	r A	sp s	Ser	D	. ~		2220 740	
	221																									,40	
		ACA Thr	GA.	T G	AA (CTT	AAT	GGT	CTT	ACA	GAA	GG	A A	CT T	AT ;	ACC	TTA	AAZ	. GC	ית ב			٠		_		
		Thr	AS]	PG.	lu I	Leu	Asn	Gly	Leu	Thr	Glu	G1	y Ti	hr T	yr 1	Chr	Leu	Lys	Al	a I'	יוו	I n	ACT	· GA	AT	2280	
																										760	
- 2	261 761	AAC Asn	GAG	G	GG (CT	TCT	ACA	GAA	ACG	CAA	TT	T AC	CG T	TA A	CT	GTA	ልሞክ	200								
	,01	Asn	Asp	• G	ly A	lla'	Ser	Thr	Glu	Thr	Gln	Ph	e Th	ır L	eu I	hr '	Val	710	Th.	A GA	AA C	AA .	agt	CC	:G	2340	
																										780	
23	41	TCT Ser	GAG	AA :	T	GT (GAC	TTT	AAT	ACA	CCT	TC:	T TC	A A	י עב	GT -	~ 1										
7	81	Ser	Glu	As	in C	ys .	Asp	Phe .	Asn	Thr	Pro	Sea	r Se	r T	ır c	Ju :	LA	GAA Ca	GA:	r TI	T G	AC .	ATT	AA	A	2400	
														- ••	0	-y 1	-eu	GIU	Ası	Ph	e A	sp :	lle	Ly	\$	800	•
24	01	AAG :	TTT	TC	TA	AC (STT :	TTT (GAG	TTA	GG b	س/ب	ר כי														
										- • • •	JUA	101	ناف	L GG	A C	LA 1	CT	TTA	AGI	` AA	TT	TA J	AA	AC	A :	2460	
										E4.										•							

501	. Ly	's Pi	ne S	er .	Asn	Va.	l Ph	e G1	u Le	u G1	y Se	r Gl	y Gl	y Pr	o Se	r Le	u Se	r As	an Le	· eu Ly	's Thr	82
2461 821	TI	T AC	TA	TT ;	LAT	TG	3 AA:	ד דר	G C2				_				•	•		•	A AAC r Asn	
2521 841	AA	C GG	T G	TA C	CT	GAT	TAT	TA	ኮ ልጥ:	A 350											r Asn A AAT s Asn	2580 860
2581 861	GC:Ala	AA1	r cc	A G	AA lu	ATA Ile	TCT Ser	Ile	AGC Ser	AAT Asn	`AGC	TTA Leu	ATT	CCT Pro	' AAT Ast	TTI Phe	GAT	GG7	GA:	TAC	TGG	2640 880
881							•					• • • •	SEL	nys	Thr	Asn	Asn	Phe	Thr	Ile	Tyr	2700 900
											-,-	non.	vai	rnr	Pro	Ser	neA	Gln	Ile	Ser		2760 920
2761 921	ATT	ACT Thr	GAI	GA As	Ť I	CT	AGT Ser	ATT Ile	AAT Asn	TTT Phe	AAG Lys	CTT Leu	TAC Tyr	CCT Pro	7.AT Asn	CCT Pro	GCT Al'a	TTA Leu	GAC Asp	GAA Glu	ACT Thr	2820 940
821	ATT	TTT	GTG	AGO	G	CT (GAA (SAT	GAA	AAA (~~~~ × ·	· ·		_				GT 2				

Figure 17d(continued)

Figure No. 180 Pyrococcus furiosus VC1(7EG1)

leader sequence: amino acids 1-24

5' ATG AGC AAG AAA AAG TTC GTC ATC GTA TCT ATC TTA ACA ATC CTT TTA GTA CAG
Met Ser Lys Lys Lys Phe Val Ile Val Ser Ile Leu Thr Ile Leu Leu Val Gln

63 72 81 90 99 108 GCA ATA TAT TTT GTA GAA AAG TAT CAT ACC TCT GAG GAC AAG TCA ACT TCA AAT Ala Ile Tyr Phe Val Glu Lys Tyr His Thr Ser Glu Asp Lys Ser Thr Ser Asn

ACC TCA TCT ACA CCA CCC CAA ACA ACA CTT TCC ACT ACC AAG GTT CTC AAG ATT Thr Ser Ser Thr Pro Pro Gln Thr Thr Leu Ser Thr Thr Lys Val Leu Lys Ile

171 180 189 198 207 216

AGA TAC CCT GAT GAC GGT GAG TGG CCA GGA GCT CCT ATT GAT AAG GAT GGT GAT

Arg Tyr Pro Asp Asp Gly Glu Trp Pro Gly Ala Pro Ile Asp Lys Asp Gly Asp

GGG AAC CCA GAA TTC TAC ATT GAA ATA AAC CTA TGG AAC ATT CTT AAT GCT ACT Gly Asn Pro Glu Phe Tyr Ile Glu Ile Asn Leu Trp Asn Ile Leu Asn Ala Thr

GGA TTT GCT GAG ATG ACG TAC AAT TTA ACC AGC GGC GTC CTT CAC TAC GTC CAA Gly Phe Ala Glu Met Thr Tyr Asn Leu Thr Ser Gly Val Leu His Tyr Val Gln

333 342 351 360 369 378 CAA CTT GAC AAC ATT GTC TTG AGG GAT AGA AGT AAT TGG GTG CAT GGA TAC CCC Gln Leu Asp Asn Ile Val Leu Arg Asp Arg Ser Asn Trp Val His Gly Tyr Pro

387 396 405 414 423 432 GAA ATA TTC TAT GGA AAC AAG CCA TGG AAT GCA AAC TAC GCA ACT GAT GGC CCA Glu Ile Phe Tyr Gly Asn Lys Pro Trp Asn Ala Asn Tyr Ala Thr Asp Gly Pro

A41 450 459 468 477 486
ATA CCA TTA CCC AGT AAA GTT TCA AAC CTA ACA GAC TTC TAT CTA ACA ATC TCC
Ile Pro Leu Pro Ser Lys Val Ser Asn Leu Thr Asp Phe Tyr Leu Thr Ile Ser

495 504 513 522 TAT AAA CTT GAG CCC AAG AAC GGC CTG CCA ATT AAC TTC GCA ATA GAA TCC TGG Tyr Lys Leu Glu Pro Lys Asn Gly Leu Pro Ile Asn Phe Ala Ile Glu Ser Trp 549 558 567 576 TTA ACG AGA GAA GCT TGG AGA ACA ACA GGA ATT AAC AGC GAT GAG CAA GAA GTA Leu Thr Arg Glu Ala Trp Arg Thr Thr Gly Ile Asn Ser Asp Glu Gln Glu Val 603 612 621 630 ATG ATA TGG ATT TAC TAT GAC GGA TTA CAA CCG GCT GGC TCC AAA GTT AAG GAG Met Ile Trp Ile Tyr Tyr Asp Gly Leu Gln Pro Ala Gly Ser Lys Val Lys Glu . 657 666 675 684 ATT GTA GTC CCA ATA ATA GTT AAC GGA ACA CCA GTA AAT GCT ACA TTT GAA GTA 693 Ile Val Val Pro Ile Ile Val Asn Gly Thr Pro Val Asn Ala Thr Phe Glu Val 711 720 729 73B TGG AAG GCA AAC ATT GGT TGG GAG TAT GTT GCA TTT AGA ATA AAG ACC CCA ATC Trp Lys Ala Asn Ile Gly Trp Glu Tyr Val Ala Phe Arg Ile Lys Thr Pro Ile 765 774 783 792 AAA GAG GGA ACA GTG ACA ATT CCA TAC GGA GCA TTT ATA AGT GTT GCA GCC AAC 801 Lys Glu Gly Thr Val Thr Ile Pro Tyr Gly Ala Phe Ile Ser Val Ala Ala Asn 819 828 837 ATT TCA AGC TTA CCA AAT TAC ACA GAA CTT TAC TTA GAG GAC GTG GAG ATT GGA Ile Ser Ser Leu Pro Asn Tyr Thr Glu Leu Tyr Leu Glu Asp Val Glu Ile Gly

873 882 891 900 909 918

ACT GAG TTT GGA ACG CCA AGC ACT ACC TCC GCC CAC CTA GAG TGG TGG ATC ACA

Thr Glu Phe Gly Thr Pro Ser Thr Thr Ser Ala His Leu Glu Trp Trp Ile Thr

927 936 945 954

AAC ATA ACA CTA ACT CCT CTA GAT AGA CCT CTT ATT TCC TAA 3'

Asn Ile Thr Leu Thr Pro Leu Asp Arg Pro Leu Ile Ser *

Figure 18b(continued)

INTERNATIONAL SEARCH REPORT

International application No. PCT/US97/22623

IPC(6)	SSIFICATION OF SUBJECT MATTER :C07H 21/04; C12N 1/20, 1/14, 5/00, 9/38, 9/42; C:435/207, 209, 252.3, 254.11, 274, 275, 320.1, 325				
	o International Patent Classification (IPC) or to both				
B. FIEL	DS SEARCHED				
	ocumentation searched (classification system followed	by classification symbols)			
U.S. :	435/207, 209, 252.3, 254.11, 274, 275, 320.1, 325;	536/23.2	· I		
Documentat	ion searched other than minimum documentation to the	extent that such documents are included	in the fields searched		
77			1		
	lata base consulted during the international search (na	me of data base and, where practicable	., search terms used)		
c. Doc	UMENTS CONSIDERED TO BE RELEVANT				
Category*	Citation of document, with indication, where ap	propriate, of the relevant passages	Relevant to claim No.		
X	GRABNITZ et al. Structure of the β	-Glucosidase Gene bglA of	1-3, 5		
	Clostridium thermocellum: Sequence Ar		species II		
Α	of Cellulases and β-Glycosidases Includ				
1	Hydrolase. Eur. J. Biochem. Septemb		4, 6-11		
	pages 301-309, see entire document.	701 1991, 101. 200, 110. 2,	', ' ' ' '		
	pages 301 307, see chare document.				
X	VOORHORST et al. Characterization	of the celB Gene Coding for	1-3, 5		
	β-Glucosidase from the Hyperthermore	_	species I and III		
A	furiosus and Its Expression and Site-Dire		species I and III		
Λ.	coli. J. Bacteriol. December 1995, Vo		4, 6-11		
		1. 177, 140. 24, pages 7103-	4, 0-11		
	7111, see entire document.				
	•				
		·			
	· · · · · ·				
Furt	her documents are listed in the continuation of Box C				
•	ocial catagories of cited documents:	"T" later document published after the int date and not in conflict with the app	lication but cited to understand		
	cument defining the general state of the art which is not considered be of particular relevance	the principle or theory underlying th			
	rlier document published on or after the international filing date	"X" document of particular relevance; the considered novel or cannot be considered.			
	ecument which may throw doubts on priority claim(s) or which is led to establish the publication date of another citation or other	when the document is taken alone			
	ecial reason (as specified)	"Y" document of particular relevance; the considered to involve an inventive	step when the document is		
	exament referring to an oral disclosure, use, exhibition or other cans	combined with one or more other suc being obvious to a person skilled in			
	ocument published prior to the international filing date but later than e priority date claimed	*&* document member of the same pater	nt family		
Date of the	actual completion of the international search	Date of mailing of the international se	earch report		
26 MAR	CH 1998	2 1 APR 1998			
	mailing address of the ISA/US	Authorized officer	· h		
	oner of Patents and Trademarks	LICA I WORDS BY S	XVX		
Washingto	on, D.C. 20231				
Facsimile N	D.C. 20231 LISA J. HOBBS, PH.D. Telephone No. (703) 308-0196				

WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6: C07H 21/04, C12N 1/20, 1/14, 5/00, 9/38, 9/42, C08B 30/04

(11) International Publication Number:

WO 98/24799

A1

(43) International Publication Date:

11 June 1998 (11.06.98)

(21) International Application Number:

PCT/US97/22623

(22) International Filing Date:

8 December 1997 (08.12.97)

(81) Designated States: AU, CA, IL, JP, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

(30) Priority Data:

60/056,916 08/949,026 6 December 1996 (06.12.96) US 10 October 1997 (10.10.97)

US

(71) Applicant (for all designated States except US): DIVERSA CORPORATION [US/US]; 10665 Sorrento Valley Road, San Diego, CA 92121 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): BYLINA, Edward, J. [US/US]; Apartment A-1, West Court, Andalusia, PA 19020 (US). SWANSON, Ronald, V. [US/US]; Apartment A, 309 No. Lemon Street, Media, PA 19063 (US). MATHUR, Eric, J. [US/US]; 2654 Galicia Way, Carlsbad, CA 92009 (US). LAM, David, E. [US/US]; 1518 West 249th Street, Harbor City, CA 90710 (US).

(74) Agent: HAILE, Lisa, A.; Fish & Richardson P.C., Suite 1400. 4225 Executive Square, La Jolla, CA 92037 (US).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: GLYCOSIDASE ENZYMES

(57) Abstract

Thermostable glycosidase enzymes derived from various Thermococcus, Staphylothermus and Pyrococcus organisms is disclosed. The enzymes are produced from native or recombinant host cells and can be utilized in the food processing industry, pharmaceutical industry and in the textile industry, detergent industry and in the baking industry.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

		DC	S	LS	Lesotho	SI	Slovenia
AL	Albania	ES	Spain	LT		SK	Slovakia
AM	Armenia	FI	Finland		Lithuania		
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
ΑU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
ΑZ	Azerbaijan	GB	United Kingdom	MC	Моласо	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR ·	Brazil	IL	Israel	MR.	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland -	MW	Malawi	US	United States of America
CA	Canada	IT	Itały	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JР	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugosłavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	zw	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	· LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

CORRECTED VERSION**

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

The state of the s							
(51) International Patent Classification ⁶ :		(11) International Publication Number: WO 98/24799					
C07H 21/04, C12N 1/20, 1/14, 5/00, 9/38, 9/42, C08B 30/04	A1	(43) International Publication Date: 11 June 1998 (11.06.98)					
(21) International Application Number: PCT/US	97/226:	LAM, David, E. [US/US]; 1518 West 249th Street, Harbor City, CA 90710 (US).					
(22) International Filing Date: 8 December 1997 (08.12.9	7) (74) Agent: HAILE, Lisa, A.; Fish & Richardson P.C., Suite 1400, 4225 Executive Square, La Jolla, CA 92037 (US).					
(30) Priority Data:							
60/056,916 6 December 1996 (06.12.96) l	JS					
08/949, 026 10 October 1997 (10.10.97)	ι	US (81) Designated States: AU, CA, IL, JP, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).					
(63) Related by Continuation (CON) or Continuation-ir (CIP) to Earlier Applications	1-Part						
US 60/056,9							
Filed on 6 December 1996 (
US Not furnish							
Filed on 10 October 1997 (10.10.9	7) claims and to be republished in the event of the receipt of amendments.					
(71) Applicant (for all designated States except US): D CORPORATION [US/US]; 10665 Somento Vall San Diego, CA 92121 (US).							
(72) Inventors; and (75) Inventors/Applicants (for US only): BYLINA, Ed [US/US]; Apartment A-1, West Court, Andalusia, I (US). SWANSON, Ronald, V. [US/US]; Apartment No. Lemon Street, Media, PA 19063 (US). MATH J. [US/US]; 2654 Galicia Way, Carlsbad, CA 920	PA 1902 nt A, 30 UR, Eri	20					

(54) Title: GLYCOSIDASE ENZYMES

(57) Abstract

Thermostable glycosidase enzymes derived from various Thermococcus, Staphylothermus and Pyrococcus organisms is disclosed. The enzymes are produced from native or recombinant host cells and can be utilized in the food processing industry, pharmaceutical industry and in the textile industry, detergent industry and in the baking industry.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

			•				
AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
ΑU	Australia	GA	Gabon	LV .	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	MŁ	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	īL	Israel	MR	Mauritania	ÜĞ	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA		lТ	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	•	KE	Кепуа	NL	Netherlands	YU	Yugoslavia
CH	_	KG	Kyrgyzstan	NO	Norway	zw	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CN	-		Republic of Korea	PL	Poland		
CN		KR	Republic of Korea	PT	Portugal		
cı		KZ	Kazakstan	RO	Romania		
Cz		LC	Saint Lucia	RU	Russian Federation		
DE	•	LI	Liechtenstein	SD	Sudan		
DH		LK	Sri Lanka	SE	Sweden		
EE		LR	Liberia	SG	Singapore		

CORRECTED VERSION.

CORRECTED VERSION **

CORRECTED VERSION***

WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)



(51) International Patent Classification 6: C07H 21/04, C12N 1/20, 1/14, 5/00, 9/38, 9/42, C08B 30/04

(11) International Publication Number:

City, CA 90710 (US).

WO 98/24799

A1

(43) International Publication Date:

11 June 1998 (11.06.98)

(21) International Application Number:

PCT/US97/22623

(22) International Filing Date:

8 December 1997 (08.12.97)

(30) Priority Data:

60/056,916

6 December 1996 (06.12.96)

US US

08/949, 026

10 October 1997 (10.10.97)

(81) Designated States: AU, CA, IL, JP, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Applications

US

60/056,916 (CIP) 6 December 1996 (06.12.96)

Filed on US

Not furnished (CIP)

Filed on

10 October 1997 (10.10.97)

(71) Applicant (for all designated States except US): DIVERSA CORPORATION [US/US]; 10665 Sorrento Valley Road, San Diego, CA 92121 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): BYLINA, Edward, J. [US/US]; Apartment A-1, West Court, Andalusia, PA 19020 (US). SWANSON, Ronald, V. [US/US]; Apartment A, 309 No. Lemon Street, Media, PA 19063 (US). MATHUR, Eric, J. [US/US]; 2654 Galicia Way, Carlsbad, CA 92009 (US). Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

LAM, David, E. [US/US]; 1518 West 249th Street, Harbor

(74) Agent: HAILE, Lisa, A.; Fish & Richardson P.C., Suite 1400, 4225 Executive Square, La Jolla, CA 92037 (US).

(54) Title: GLYCOSIDASE ENZYMES

(57) Abstract

Thermostable glycosidase enzymes derived from various Thermococcus, Staphylothermus and Pyrococcus organisms is disclosed. The enzymes are produced from native or recombinant host cells and can be utilized in the food processing industry, pharmaceutical industry and in the textile industry, detergent industry and in the baking industry.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

			•				
AL A	lbania	ES	Spain	LS	Lesotho	SI	Slovenia
AM A	rmenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT A	ustria	FR	France	LU	Luxembourg	SN	Senegal
AU A	ustralia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ A:	zerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA Bo	osnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB Ba	arbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE Be	elgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF B	urkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG B	ulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ B	enin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR B	razil	IL	Israel	MR	Mauritania	UG	Uganda
BY B	elarus	IS	Iceland	MW	Malawi	US	United States of America
CA C	anada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF C	entral African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG C	ongo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH S	witzerland	KG	Kyrgyzstan	NO	Norway	zw	Zimbabwe
CI C	ôte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM C	ameroon		Republic of Korea	PL	Poland		
CN C	'hina	KR	Republic of Korea	PT	Portugal		
cu c	luba	KZ	Kazakstan	RO	Romania		
CZ C	zech Republic	LC	Saint Lucia	RU	Russian Federation		
DE G	ermany	LI	Liechtenstein	SD	Sudan		
DK D	enmark	LK	Sri Lanka	SE	Sweden		
EE E	stonia	LR	Liberia	SG	Singapore		
EE E	stonia	LR	Liberia	SG	Singapore		

GLYCOSIDASE ENZYMES

BACKGROUND OF THE INVENTION

1. Field of the Inventions

This invention relates to newly identified polynucleotides, polypeptides encoded by such polynucleotides, the use of such polynucleotides and polypeptides, as well as the production and isolation of such polynucleotides and polypeptides. More particularly, the polynucleotides and polypeptides of the present invention has been putatively identified as glucosidases, α -galactosidases, β -galactosidases, β -mannosidases, β -mannosidases, and pullalanases.

10 2. Description of Related Art

The glycosidic bond of β -galactosides can be cleaved by different classes of enzymes: (i) phospho- β -galactosidases (EC3.2.1.85) are specific for a phosphorylated substrate generated via phosphoenolpyruvate phosphotransferase system (PTS)-dependent uptake; (ii) typical β -galactosidases (EC 3.2.1.23), represented by the *Escherichia coli* LacZ

- enzyme, which are relatively specific for β-galactosides; and (iii) β-glucosidases (EC 3.2.1.21) such as the enzymes of Agrobacterium faecalis, Clostridium thermocellum, Pyrococcus furiosus or Sulfolobus solfataricus (Day, A.G. and Withers, S.G., (1986) Purification and characterization of a β-glucosidase from Alcaligenes faecalis. Can. J. Biochem. Cell. Biol. 64, 914-922; Kengen, S.W.M., et al. (1993) Eur. J. Biochem., 213,
- 305-312; Ait, N., Cruezet, N. and Cattaneo, J. (1982) Properties of β-glucosidase purified from Clostridium thermocellum. J. Gen. Microbiol. 128, 569-577; Grogan, D.W. (1991) Evidence that β-galactosidase of Sulfolobus solfataricus is only one of several activities of a thermostable β-D-glycodiase. Appl. Environ. Microbiol. 57, 1644-1649). Members of the latter group, although highly specific with respect to the β-anomeric configuration
- of the glycosidic linkage, often display a rather relaxed substrate specificity and hydrolyze β -glucosides as well as β -fucosides and β -galactosides.

Generally, α -galactosidases are enzymes that catalyze the hydrolysis of galactose groups on a polysaccharide backbone or hydrolyze the cleavage of di- or oligosaccharides comprising galactose.

Generally, \(\beta\)-mannanases are enzymes that catalyze the hydrolysis of mannose groups internally on a polysaccharide backbone or hydrolyze the cleavage of di- or oligosaccaharides comprising mannose groups. \(\beta\)-mannosidases hydrolyze non-reducing, terminal mannose residues on a mannose-containing polysaccharide and the cleavage of di- or oligosaccaharides comprising mannose groups.

Guar gum is a branched galactomannan polysaccharide composed of β-1,4 linked mannose backbone with α-1,6 linked galactose side chains. The enzymes required for the degradation of guar are β-mannanase, β-mannosidase and α-galactosidase. β-mannanase hydrolyses the mannose backbone internally and β-mannosidase hydrolyses non-reducing, terminal mannose residues. α-galactosidase hydrolyses α-linked galactose groups.

Galactomannan polysaccharides and the enzymes that degrade them have a variety of applications. Guar is commonly used as a thickening agent in food and is utilized in hydraulic fracturing in oil and gas recovery. Consequently, galactomannanases are industrially relevant for the degradation and modification of guar. Furthermore, a need exists for thermostable galactomannases that are active in extreme conditions associated with drilling and well stimulation.

There are other applications for these enzymes in various industries, such as in the beet sugar industry. 20-30% of the domestic U.S. sucrose consumption is sucrose from sugar beets. Raw beet sugar can contain a small amount of raffinose when the sugar beets are stored before processing and rotting begins to set in. Raffinose inhibits the crystallization of sucrose and also constitutes a hidden quantity of sucrose. Thus, there is merit to eliminating raffinose from raw beet sugar. α-Galactosidase has also been used

as a digestive aid to break down raffinose, stachyose, and verbascose in such foods as beans and other gassy foods.

 β -galactosidases which are active and stable at high temperatures appear to be superior enzymes for the production of lactose-free dietary milk products (Chaplin, M.F. and 5 Bucke, C. (1990) In: Enzyme Technology, pp. 159-160, Cambridge University Press, Cambridge, UK). Also, several studies have demonstrated the applicability of β galactosidases to the enzymatic synthesis of oligosaccharides via transglycosylation reactions (Nilsson, K.G.I. (1988) Enzymatic synthesis of oligosaccharides. Trends Biotechnol. 6, 156-264; Cote, G.L. and Tao, B.Y. (1990) Oligosaccharide synthesis by enzymatic transglycosylation. Glycoconjugate J. 7, 145-162). Despite the commercial potential, only a few β-galactosidases of thermophiles have been characterized so far. Two genes reported are β-galactoside-cleaving enzymes of the hyperthermophilic bacterium Thermotoga maritima, one of the most thermophilic organotrophic eubacteria described to date (Huber, R., Langworthy, T.A., König, H., Thomm, M., Woese, C.R., Sleytr, U.B. and Stetter, K.O. (1986) T. martima sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90°C, Arch. Microbiol. 144, 324-333) one of the most thermophilic organotrophic eubacteria described to date. The gene products have been identified as a β-galactosidase and a β-glucosidase.

Pullulanase is well known as a debranching enzyme of pullulan and starch. The enzyme hydrolyzes α-1,6-glucosidic linkages on these polymers. Starch degradation for the production or sweeteners (glucose or maltose) is a very important industrial application of this enzyme. The degradation of starch is developed in two stages. The first stage involves the liquefaction of the substrate with α-amylase, and the second stage, or saccharification stage, is performed by β-amylase with pullalanase added as a debranching enzyme, to obtain better yields.

Endoglucanases can be used in a variety of industrial applications. For instance, the endoglucanases of the present invention can hydrolyze the internal \(\beta -1,4-glycosidic \)

bonds in cellulose, which may be used for the conversion of plant biomass into fuels and chemicals. Endoglucanases also have applications in detergent formulations, the textile industry, in animal feed, in waste treatment, and in the fruit juice and brewing industry for the clarification and extraction of juices.

Brief Description of the Drawings

The following drawings are illustrative of embodiments of the invention and are not meant to limit the scope of the invention as encompassed by the claims.

Figures 1a-b are the full-length DNA and corresponding deduced amino acid sequence of M11TL of the present invention. Sequencing was performed using a 378 automated DNA sequencer for all sequences of the present invention (Applied Biosystems, Inc.).

Figure 2 is an illustration of the full-length DNA and corresponding deduced amino acid sequence of OC1/4V-33B/G.

Figure 3 is an illustration of the full-length DNA and corresponding deduced amino acid sequence of F1-12G.

Figures 4a-b are the full-length DNA and corresponding deduced amino acid sequence of 9N2-31B/G.

Figures 5a-b are the full-length DNA and corresponding deduced amino acid sequence of MSB8-6G.

15 Figure 6 is the full-length DNA and corresponding deduced amino acid sequence of AEDII12RA-18B/G.

Figures 7a-b are the full-length DNA and corresponding deduced amino acid sequence of GC74-22G.

Figures 8a-b are the full-length DNA and corresponding deduced amino acid sequence of VC1-7G1.

Figures 9a-c are the full-length DNA and corresponding deduced amino acid sequence of 37GP1.

Figures 10a-c are the full-length DNA and corresponding deduced amino acid sequence of 6GC2.

5 Figures 11a-d are the full-length DNA and corresponding deduced amino acid sequence of 6GP2.

Figures 12a-c are the full-length DNA and corresponding deduced amino acid sequence of 63GB1.

Figures 13a-b are the full-length DNA and corresponding deduced amino acid sequence of OC1/4V.

Figures 14a-e are the full-length DNA and corresponding deduced amino acid sequence of 6GP3.

Figures 15a-d are the full-length DNA and corresponding deduced amino acid sequence of *Thermotoga maritima* MSB8-6GP2.

15 Figures 16a-c are the full-length DNA and corresponding deduced amino acid sequence of *Thermotoga maritima* MSB8-6GB4.

Figures 17a-d are the full-length DNA and corresponding deduced amino acid sequence of *Banki gouldi* 37GP4.

Figures 18a-b are the full-length DNA and corresponding deduced amino acid sequence of *Pyrococcus furiosus* VC1-7EG1.

SUMMARY OF THE INVENTION

In a preferred embodiment of the present invention, there are provided isolated nucleic acids (polynucleotides) which encode mature enzymes having the deduced amino acid sequences of Figures 1-18 (SEQ ID NOS: 15-28 and 61-64).

In another embodiment, the invention provides a method for producing a polypeptide including culturing host cells containing the polynucleotide of Figures 1-18 and expressing from the host cell a polypeptide encoded by the polynucleotide and isolating the polypeptide.

In another embodiment, the invention provides an enzyme selected from the group consisting of an enzyme having an amino acid sequence set forth in SEQ ID NOS: 15-28 or 61-64 and an enzyme which has at least 30 consecutive amino acid residue as an enzyme having an amino acid sequence set forth in SEQ ID NOS: 15-28 or 61-64.

In yet another embodiment, the invention provides a method for generating glucose from soluble cell oligosaccharides which includes contacting a sample containing oligosaccharides with an effective amount of an enzyme selected from the group of enzymes having the amino acid sequence set forth in SEQ ID NOS: 15-28, 61-63 and 64 such that glucose is produced

The publications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

Definitions

"Monosaccharide", as used herein, refers to a single polyhydroxy aldehyde or ketone unit.

"Oligosaccharide", as used herein, consist of short chains of monosaccharide units joined together by covalent bonds. Of these, the most abundant are the disaccharides, which have two monosaccharide units.

"Polysaccharide", as used herein, consists of long chains having many monosaccharide units.

The term "gene" means the segment of DNA involved in producing a polypeptide chain; it includes regions preceding and following the coding region (leader and trailer) as well as intervening sequences (introns) between individual coding segments (exons).

A coding sequence is "operably linked to" another coding sequence when RNA polymerase will transcribe the two coding sequences into a single mRNA, which is then translated into a single polypeptide having amino acids derived from both coding sequences. The coding sequences need not be contiguous to one another so long as the expressed sequences ultimately process to produce the desired protein.

"Recombinant" enzymes refer to enzymes produced by recombinant DNA techniques; 5 *i.e.*, produced from cells transformed by an exogenous DNA construct encoding the desired enzyme. "Synthetic" enzymes are those prepared by chemical synthesis.

A DNA "coding sequence of" or a "nucleotide sequence encoding" a particular enzyme, is a DNA sequence which is transcribed and translated into an enzyme when placed under the control of appropriate regulatory sequences.

20 <u>Detailed Description of the Invention</u>

The polynucleotides and polypeptides of the present invention have been identified as glucosidases, α -galactosidases, β -galactosidases, β -mannosidases, β -mannanases, endoglucanases, and pullalanases as a result of their enzymatic activity.

In accordance with one aspect of the present invention, there are provided novel enzymes, as well as active fragments, analogs and derivatives thereof.

In accordance with another aspect of the present invention, there are provided isolated nucleic acid molecules encoding the enzymes of the present invention including mRNAs, cDNAs, genomic DNAs as well as active analogs and fragments of such enzymes.

In accordance with yet a further aspect of the present invention, there is provided a process for producing such polypeptides by recombinant techniques comprising culturing recombinant prokaryotic and/or eukaryotic host cells, containing a nucleic acid sequence of the present invention, under conditions promoting expression of said enzymes and subsequent recovery of said enzymes.

In accordance with yet a further aspect of the present invention, there is provided a process for utilizing such enzymes, or polynucleotides encoding such enzymes for hydrolyzing lactose to galactose and glucose for use in the food processing industry, the pharmaceutical industry, for example, to treat intolerance to lactose, as a diagnostic reporter molecule, in corn wet milling, in the fruit juice industry, in baking, in the textile industry and in the detergent industry.

In accordance with yet a further aspect of the present invention, there is provided a process for utilizing such enzymes for hydrolyzing guar gum (a galactomannan polysaccharide) to remove non-reducing terminal mannose residues. Further polysaccharides such as galactomannan and the enzymes according to the invention that degrade them have a variety of applications. Guar gum is commonly used as a thickening agent in food and also is utilized in hydraulic fracturing in oil and gas recovery. Consequently, mannanases are industrially relevant for the degradation and modification of guar gums. Furthermore, a need exists for thermostable mannases that are active in extreme conditions associated with drilling and well stimulation.

In accordance with yet a further aspect of the present invention, there are also provided nucleic acid probes comprising nucleic acid molecules of sufficient length to specifically hybridize to a nucleic acid sequence of the present invention.

In accordance with yet a further aspect of the present invention, there is provided a process for utilizing such enzymes, or polynucleotides encoding such enzymes, for *in vitro* purposes related to scientific research, for example, to generate probes for identifying similar sequences which might encode similar enzymes from other organisms by using certain regions, *i.e.*, conserved sequence regions, of the nucleotide sequence.

These and other aspects of the present invention should be apparent to those skilled in the art from the teachings herein.

The polynucleotides of this invention were originally recovered from genomic gene libraries derived from the following organisms:

M11TL is a new species of *Desulfurococcus* isolated from Diamond Pool in Yellowstone National Park. The organism grows optimally at 85-88°C, pH 7.0 in a low salt medium containing yeast extract, peptone, and gelatin as substrates with a N₂/CO₂ gas phase.

OC1/4V is from the genus *Thermotoga*. The organism was isolated from Yellowstone National Park. It grows optimally at 75°C in a low salt medium with cellulose as a substrate and N_2 in gas phase.

Pyrococcus furiosus VC1 and (7EG1) is from the genus Pyrococcus. VC1 was isolated from Vulcano, Italy. It grows optimally at 100°C in a high salt medium (marine) containing elemental sulfur, yeast extract, peptone and starch as substrates and N₂ in gas phase.

Staphylothermus marinus F1 is a from the genus Staphylothermus. F1 was isolated from Vulcano, Italy. It grows optimally at 85°C, pH 6.5 in high salt medium (marine) containing elemental sulfur and yeast extract as substrates and N₂ in gas phase.

Thermococcus 9N-2 is from the genus Thermococcus 9N-2 was isolated from diffuse vent fluid in the East Pacific Rise. It is a strict anaerobe that grows optimally at 87°C.

Thermotoga maritima MSB8 and MSB8 (Clone # 6GP2 and 6GB4) is from the genus Thermotogo, and was isolated from Vulcano, Italy. MSB8 grows optimally at 85°C, pH 6.5 in a high salt medium (marine) containing starch and yeast extract as substrates and N₂ in gas phase.

Thermococcus alcaliphilus AEDII12RA is from the genus Thermococcus. AEDII12RA grows optimally at 85°C, pH 9.5 in a high salt medium (marine) containing polysulfides and yeast extract as substrates and N₂ in gas phase.

Thermococcus chitonophagus GC74 is from the genus Thermococcus. GC74 grows optimally at 85°C, pH 6.0 in a high salt medium (marine) containing chitin, meat extract, elemental sulfur and yeast extract as substrates and N₂ in gas phase. AEPII 1a grows optimally at 85°C at pH 6.5 in marine medium under anaerobic conditions. It has many substrates. Bankia gouldi is from the genus Bankia.

Accordingly, the polynucleotides and enzymes encoded thereby are identified by the organism from which they were isolated, and are sometimes hereinafter referred to as "M11TL" (Figure 1 and SEQ ID NOS:1 and 15), "OC1/4V-33B/G" (Figure 2 and SEQ ID NOS:2 and 16), "F1-12G" (Figure 3 and SEQ ID NOS:3 and 17), "9N2-31B/G" (Figure 4 and SEQ ID NOS:4 and 18), "MSB8" (Figure 5 and SEQ ID NOS:5 and 19), "AEDII12RA-18B/G" (Figure 6 and SEQ ID NOS:6 and 20), "GC74-22G" (Figure 7 and SEQ ID NOS:7 and 21), "VC1-7G1" (Figure 8 and SEQ ID NOS:8 and 22), "37GP1" (Figure 9 and SEQ ID NOS: 9 and 23), "6GC2" (Figure 10 and SEQ ID NOS: 10 and

24), "6GP2" (Figure 11 and SEQ ID NOS:11 and 25), "AEPII 1a" (Figure 12 and SEQ ID NOS:12 and 26), "OC1/4V" (Figure 13 and SEQ ID NOS:13 and 27), and "6GP3" (Figure 14 and SEQ ID NOS:28), "MSB8-6GP2" (Figure 15 and SEQ ID NOS:57 and 61), "MSB8-6GB4"(Figure 16 and SEQ ID NOS:58 and 62), "VC1-7EG1"(Figure 17 and SEQ ID NOS:59 and 63), and 37GP4 (Figure 18 and SEQ ID NOS:60 and 64).

The polynucleotides and polypeptides of the present invention show identity at the nucleotide and protein level to known genes and proteins encoded thereby as shown in Table 1.

Table 1

10	Clone		Protein Identity	Nucleic Acid Identity
-	M11TL-29G	Sulfolobus sulfataricus DSM 1616/P1, β- galactosidase	51%	55%
	OC1/4V-33B/G	Caldocellum saccharolyticum, β-glucosidase	52%	57%
	Staphylothermus marinus F1-12G	Bacillus polymyxa, β- galactosidase	36%	48%
15	Thermococcus 9N2- 31B/G	Sulfolobus sulfataricus ATCC 49255/MT4, β- galactosidase	51%	50%
	Thermotoga maritima MSB8-6G	Clostridium thermocellum	45%	53%
20	Thermococcus AEDII12RA-18B/G	Bacillus polymyxa, β-galactosidase	34%	48%

1	,			
1	Thermococcus chitonophagus GC74- 22G	Sulfolobus sulfataricus ATCC 49255/MT4, β- galactosidase	46%	54%
5	Pyrococcus furiosus VC1-7G1	Sulfolobus sulfataricus/MT-4 β- galactosidase	46.4%	52.5%
	Thermotoga maritima α-galactosidase (6GC2)	Pediococcus pentosaceaus α-galactosidase	49%	29%
10	Thermotoga maritima B-mannanase (6GP2)	Aspergillus aculeatus mannanase	56%	37%
	AEPII 1a ß- mannosidase (63GB1)	Sulfolobus solfactaricus ß- galactosidase	78%	56%
15	OC1/4V endoglucanase (33GP1)	Clostridium thermocellum endo-1,4-ß-endoglucanase	65%	43%
	Thermotoga mariti@aldo pullalanase (6GP3)	cellum saccharolyticum α- destrom 6 glucanohydralase	72	53
	Bankia gouldi mix Endoglucanase	None available		
20	(37GP1)			

The polynucleotides and enzymes of the present invention show homology to each other as shown in Table 2.

5

Table 2

Clone	Gene/Protein with Closest Homology	Protein Identity	Nucleic Acid Identity
Staphylothermus marinus F1-12G	Thermococcus AEDII12RA-18B/G, β- galactosidase, glucosidase	55%	57%
Thermococcus 9N2- 31B/G	Thermococcus chitonophagus GC74- 22G-glucosidase`	74%	66%
Pyrococcus furiosus VC1-7G1	Pyrococcus furiosus VC1- 7B/G β-galactosidase	46.4%	54%

All the clones identified in Tables 1 and 2 encode polypeptides which have α -glycosidase 10 or β -glycosidase activity.

This invention, in addition to the isolated nucleic acid molecules encoding the enzymes of the present invention, also provide substantially similar sequences. Isolated nucleic acid sequences are substantially similar if: (i) they are capable of hybridizing under conditions hereinafter described, to the polynucleotides of SEQ ID NOS: 1-14 and 57-60; (ii) or they encode DNA sequences which are degenerate to the polynucleotides of SEQ ID NOS: 1-14 and 57-60. Degenerate DNA sequences encode the amino acid sequences of SEQ ID NOS:15-28 and 61-64, but have variations in the nucleotide coding sequences. As used herein, substantially similar refers to the sequences having similar identity to the sequences of the instant invention. The nucleotide sequences that are substantially the same can be identified by hybridization or by sequence comparison. Enzyme sequences that are substantially the same can be identified by one or more of the

following: proteolytic digestion, gel electrophoresis and/or microsequencing.

One means for isolating the nucleic acid molecules encoding the enzymes of the present invention is to probe a gene library with a natural or artificially designed probe using art recognized procedures (see, for example: Current Protocols in Molecular Biology, 5 Ausubel F.M. et al. (EDS.) Green Publishing Company Assoc. and John Wiley Interscience, New York, 1989, 1992). It is appreciated to one skilled in the art that the polynucleotides of SEQ ID NOS: 1-14 and 57-60 or fragments thereof (comprising at least 12 contiguous nucleotides), are particularly useful probes. Other particular useful probes for this purpose are hybridizable fragments to the sequences of SEQ ID NOS: 1-14 and 57-60 (i.e., comprising at least 12 contiguous nucleotides).

With respect to nucleic acid sequences which hybridize to specific nucleic acid sequences disclosed herein, hybridization may be carried out under conditions of reduced stringency, medium stringency or even stringent conditions. As an example of oligonucleotide hybridization, a polymer membrane containing immobilized denatured nucleic acids is first prehybridized for 30 minutes at 45°C in a solution consisting of 0.9 M NaCl, 50 mM NaH₂PO₄, pH 7.0, 5.0 mM Na₂EDTA, 0.5% SDS, 10X Denhardt's, and 0.5 mg/ml polyriboadenylic acid. Approximately 2 X 10⁷ cpm (specific activity 4-9 X 10⁸ cpm/ug) of P end-labeled oligonucleotide probe are then added to the solution. After 12-16 hours of incubation, the membrane is washed for 30 minutes at room temperature in 1X SET (150 mM NaCl, 20 mM Tris hydrochloride, pH 7.8, 1 mM Na₂EDTA) containing 0.5% SDS, followed by a 30 minute wash in fresh 1X SET at Tm 10°C for the oligonucleotide probe. The membrane is then exposed to auto-radiographic film for detection of hybridization signals.

Stringent conditions means hybridization will occur only if there is at least 90% identity, preferably at least 95% identity and most preferably at least 97% identity between the sequences. Further, it is understood that a section of a 100 bps sequence that is 95 bps in length has 95% identity with the 1090 bps sequence from which it is obtained. See J.

Sambrook et al., Molecular Cloning, A Laboratory Manual, 2d Ed., Cold Spring Harbor Laboratory (1989) which is hereby incorporated by reference in its entirety. Also, it is understood that a fragment of a 100 bps sequence that is 95 bps in length has 95% identity with the 100 bps sequence from which it is obtained.

- As used herein, a first DNA (RNA) sequence is at least 70% and preferably at least 80% identical to another DNA (RNA) sequence if there is at least 70% and preferably at least a 80% or 90% identity, respectively, between the bases of the first sequence and the bases of the another sequence, when properly aligned with each other, for example when aligned by BLASTN.
- "Identity" as the term is used herein, refers to a polynucleotide sequence which comprises a percentage of the same bases as a reference polynucleotide (SEQ ID NOS:1-14 and 57-60). For example, a polynucleotide which is at least 90% identical to a reference polynucleotide, has polynucleotide bases which are identical in 90% of the bases which make up the reference polynucleotide and may have different bases in 10% of the bases which comprise that polynucleotide sequence.

The present invention relates polynucleotides which differ from the reference polynucleotide such that the changes are silent changes, for example the change do not alter the amino acid sequence encoded by the polynucleotide. The present invention also relates to nucleotide changes which result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference polynucleotide. In a preferred aspect of the invention these polypeptides retain the same biological action as the polypeptide encoded by the reference polynucleotide.

It is also appreciated that such probes can be and are preferably labeled with an analytically detectable reagent to facilitate identification of the probe. Useful reagents include but are not limited to radioactivity, fluorescent dyes or enzymes capable of catalyzing the formation of a detectable product. The probes are thus useful to isolate

complementary copies of DNA from other sources or to screen such sources for related sequences.

The polynucleotides of this invention were recovered from genomic gene libraries from the organisms listed in Table 1. For example, gene libraries can be generated in the Lambda ZAP II cloning vector (Stratagene Cloning Systems). Mass excisions can be performed on these libraries to generate libraries in the pBluescript phagemid. Libraries are thus generated and excisions performed according to the protocols/methods hereinafter described.

The excision libraries are introduced into the *E. coli* strain BW14893 F'kan1A.

Expression clones are then identified using a high temperature filter assay. Expression clones encoding several glucanases and several other glycosidases are identified and repurified. The polynucleotides, and enzymes encoded thereby, of the present invention, yield the activities as described above.

The coding sequences for the enzymes of the present invention were identified by screening the genomic DNAs prepared for the clones having glucosidase or galactosidase activity.

An example of such an assay is a high temperature filter assay wherein expression clones were identified by use of high temperature filter assays using buffer Z (see recipe below) containing 1 mg/ml of the substrate 5-bromo-4-chloro-3-indolyl-β-D-glucopyranoside (XGLU) (Diagnostic Chemicals Limited or Sigma) after introducing an excision library into the *E. coli* strain BW14893 F'kan1A. Expression clones encoding XGLUases were identified and repurified from M11TL, OC1/4V, Pyrococcus furiosus VC1, Staphylothemus marinus F1, Thermococcus 9N-2, Thermotoga maritima MSB8, Thermococcus alcaliphilus AEDII12RA, and Thermococcus chitonophagus GC74.

<u>Z-buffer:</u> (referenced in Miller, J.H. (1992) A Short Course in Bacterial Genetics, p. 445.)

per liter:

5

15

Na₂HPO₄-7H₂O 16.1g NaH₂PO₄-7H₂O 5.5g KCl 0.75g MgSO₄-7H₂O 0.246g β -mercaptoethanol 2.7ml

Adjust pH to 7.0

10 High Temperature Filter Assay

(1) The f factor f'kan (from E. coli strain CSH118)(1) was introduced into the pho-phh-lac-strain BW14893(2). BW13893(2). The filamentous phage library was plated on the resulting strain, BW14893 F'kan. (Miller, J.H. (1992) A Short Course in Bacterial Genetics; Lee, K.S., Metcalf, et al., (1992) Evidence for two phosphonate degradative pathways in Enterobacter Aerogenes, J. Bacteriol., 174:2501-2510.

- (2) After growth on 100 mm LB plates containing 100 μg/ml ampicillin, 80 μg/ml nethicillin and 1mM IPTG, colony lifts were performed using Millipore HATF membrane filters.
- The colonies transferred to the filters were lysed with chloroform vapor in 150 mm glass petri dishes.
 - (4) The filters were transferred to 100 mm glass petri dishes containing a piece of Whatman 3MM filter paper saturated with buffer.
 - (a) when testing for galactosidase activity (XGALase), 3MM paper was saturated with Z buffer containing 1 mg/ml XGAL (ChemBridge Corporation). After transferring filter bearing lysed colonies to the glass petri dish, placed dish in oven at 80-85°C.

25

(b) when testing for glucosidase (XGLUase), 3MM paper was saturated with Z buffer containing 1 mg/ml XGLU. After transferring filter bearing lysed colonies to the glass petri dish, placed dish in oven at 80-85°C.

'Positives' were observed as blue spots on the filter membranes. Used the 5 (5) following filter rescue technique to retrieve plasmid from lysed positive colony. Used pasteur pipette (or glass capillary tube) to core blue spots on the filter membrane. Placed the small filter disk in an Eppendorf tube containing 20 µl water. Incubated the Eppendorf tube at 75°C for 5 minutes 10 followed by vortexing to elute plasmid DNA off filter. This DNA was transformed into electrocompetent E. coli cells DH10B for Thermatoga maritima MSB8-6G, Staphylothermus marinus F1-12G, Thermococcus AEDII12RA-18B/G, Thermococcus chitonophagus GC74-22G, M11Tl and OC1/4V. Electrocompetent BW14893 F'kan1A E. coli were used for 15 Thermococcus 9N2-31B/G, and Pyrococcus furiosus VC1-7G1. Repeated filter-lift assay on transformation plates to identify 'positives'. Return transformation plates to 37°C incubator after filter lift to regenerate colonies. Inoculate 3 ml LB liquid containing 100 µg/ml ampicillin with repurified positives and incubate at 37°C overnight. Isolate plasmid DNA from these 20 cultures and sequence plasmid insert. In some instances where the plates used for the initial colony lifts contained non-confluent colonies, a specific colony corresponding to a blue spot on the filter could be identified on a regenerated plate and repurified directly, instead of using the filter rescue technique.

Another example of such an assay is a variation of the high temperature filter assay wherein colony-laden filters are heat-killed at different temperatures (for example, 105°C for 20 minutes) to monitor thermostability. The 3MM paper is saturated with different buffers (i.e., 100 mM NaCl, 5 mM MgCl₂, 100 mM Tris-Cl (pH 9.5)) to determine enzyme activity under different buffer conditions.

A β-glucosidase assay may also be employed, wherein GlcpβNp is used as an artificial substrate (aryl-β-glucosidase). The increase in absorbance at 405 nm as a result of p-nitrophenol (pNp) liberation was followed on a Hitachi U-1100 spectrophotometer, equipped with a thermostatted cuvette holder. The assays may be performed at 80°C or 90°C in closed 1-ml quartz cuvette. A standard reaction mixture contains 150 mM trisodium substrate, pH 5.0 (at 80°C), and 0.95 mM pNp derivative pNp = 0.561 mM⁻¹ cm⁻¹). The reaction mixture is allowed to reach the desired temperature, after which the reaction is started by injecting an appropriate amount of enzyme (1.06 ml final volume).

1 U β -glucosidase activity is defined as that amount required to catalyze the formation of 1.0 μ mol pNp/min. D-cellobiose may also be used as a substrate.

An ONPG assay for β -galactosidase activity is described by Miller, J.H. (1992) A Short Course in Bacterial Genetics and Mill, J.H. (1992) Experiments in Molecular Genetics, the contents of which are hereby incorporated by reference in their entirety.

A quantitative fluorometric assay for β-galactosidase specific activity is described by:

Youngman P., (1987) Plasmid Vectors for Recovering and Exploiting Tn917

Transpositions in Bacillus and other Gram-Positive Bacteria. In Plasmids: A Practical approach (ed. K. Hardy) pp 79-103. IRL Press, Oxford. A description of the procedure can be found in Miller (1992) p. 75-77, the contents of which are incorporated by reference herein in their entirety.

The polynucleotides of the present invention may be in the form of DNA which DNA includes cDNA, genomic DNA, and synthetic DNA. The DNA may be double-stranded or single-stranded, and if single stranded may be the coding strand or non-coding (antisense) strand. The coding sequences which encodes the mature enzymes may be identical to the coding sequences shown in Figures 1-8 (SEQ ID NOS: 1-14 and 57-60)

or may be a different coding sequence which coding sequence, as a result of the redundancy or degeneracy of the genetic code, encodes the same mature enzymes as the DNA of Figures 1-18 (SEQ ID NOS: 1-14 and 57-60).

The polynucleotide which encodes for the mature enzyme of Figures 1-18 (SEQ ID NOS: 15-28 and 61-64) may include, but is not limited to: only the coding sequence for the mature enzyme; the coding sequence for the mature enzyme and additional coding sequence such as a leader sequence or a proprotein sequence; the coding sequence for the mature enzyme (and optionally additional coding sequence) and non-coding sequence, such as introns or non-coding sequence 5' and/or 3' of the coding sequence for the mature enzyme.

Thus, the term "polynucleotide encoding an enzyme (protein)" encompasses a polynucleotide which includes only coding sequence for the enzyme as well as a polynucleotide which includes additional coding and/or non-coding sequence.

The present invention further relates to variants of the hereinabove described polynucleotides which encode for fragments, analogs and derivatives of the enzymes having the deduced amino acid sequences of Figures 1-18 (SEQ ID NOS: 15-28 and 61-64). The variant of the polynucleotide may be a naturally occurring allelic variant of the polynucleotide or a non-naturally occurring variant of the polynucleotide.

Thus, the present invention includes polynucleotides encoding the same mature enzymes as shown in Figures 1-18 (SEQ ID NOS: 15-28 and 61-64) as well as variants of such polynucleotides which variants encode for a fragment, derivative or analog of the enzymes of Figures 1-18 (SEQ ID NOS: 15-28 and 61-64). Such nucleotide variants include deletion variants, substitution variants and addition or insertion variants.

As hereinabove indicated, the polynucleotides may have a coding sequence which is a naturally occurring allelic variant of the coding sequences shown in Figures 1-18 (SEQ

ID NOS: 1-14 and 57-60). As known in the art, an allelic variant is an alternate form of a polynucleotide sequence which may have a substitution, deletion or addition of one or more nucleotides, which does not substantially alter the function of the encoded enzyme.

Fragments of the full length gene of the present invention may be used as a hybridization probe for a cDNA or a genomic library to isolate the full length DNA and to isolate other DNAs which have a high sequence similarity to the gene or similar biological activity. Probes of this type preferably have at least 10, preferably at least 15, and even more preferably at least 30 bases and may contain, for example, at least 50 or more bases. The probe may also be used to identify a DNA clone corresponding to a full length transcript and a genomic clone or clones that contain the complete gene including regulatory and promotor regions, exons, and introns. An example of a screen comprises isolating the coding region of the gene by using the known DNA sequence to synthesize an oligonucleotide probe. Labeled oligonucleotides having a sequence complementary to that of the gene of the present invention are used to screen a library of genomic DNA to determine which members of the library the probe hybridizes to.

The present invention further relates to polynucleotides which hybridize to the hereinabove-described sequences if there is at least 70%, preferably at least 90%, and more preferably at least 95% identity between the sequences. The present invention particularly relates to polynucleotides which hybridize under stringent conditions to the hereinabove-described polynucleotides. As herein used, the term "stringent conditions" means hybridization will occur only if there is at least 95% and preferably at least 97% identity between the sequences. The polynucleotides which hybridize to the hereinabove described polynucleotides in a preferred embodiment encode enzymes which either retain substantially the same biological function or activity as the mature enzyme encoded by the DNA of Figures 1-18 (SEQ ID NOS: 1-14 and 57-60).

Alternatively, the polynucleotide may have at least 15 bases, preferably at least 30 bases, and more preferably at least 50 bases which hybridize to any part of a polynucleotide of

the present invention and which has an identity thereto, as hereinabove described, and which may or may not retain activity. For example, such polynucleotides may be employed as probes for the polynucleotides of SEQ ID NOS: 1-14 and 57-60, for example, for recovery of the polynucleotide or as a diagnostic probe or as a PCR primer.

Thus, the present invention is directed to polynucleotides having at least a 70% identity, preferably at least 90% identity and more preferably at least a 95% identity to a polynucleotide which encodes the enzymes of SEQ ID NOS: 15-28 and 61-64 as well as fragments thereof, which fragments have at least 15 bases, preferably at least 30 bases and most preferably at least 50 bases, which fragments are at least 90% identical, preferably at least 95% identical and most preferably at least 97% identical under stringent conditions to any portion of a polynucleotide of the present invention.

The present invention further relates to enzymes which have the deduced amino acid sequences of Figures 1-18 (SEQ ID NOS: 15-28 and 61-64) as well as fragments, analogs and derivatives of such enzyme.

- The terms "fragment," "derivative" and "analog" when referring to the enzymes of Figures 1-18 (SEQ ID NOS: 15-28 and 61-64) means enzymes which retain essentially the same biological function or activity as such enzymes. Thus, an analog includes a proprotein which can be activated by cleavage of the proprotein portion to produce an active mature enzyme.
- 20 The enzymes of the present invention may be a recombinant enzyme, a natural enzyme or a synthetic enzyme, preferably a recombinant enzyme.
 - The fragment, derivative or analog of the enzymes of Figures 1-18 (SEQ ID NOS: 15-28 and 61-64) may be (i) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not

be one encoded by the genetic code, or (ii) one in which one or more of the amino acid residues includes a substituent group, or (iii) one in which the mature enzyme is fused with another compound, such as a compound to increase the half-life of the enzyme (for example, polyethylene glycol), or (iv) one in which the additional amino acids are fused to the mature enzyme, such as a leader or secretory sequence or a sequence which is employed for purification of the mature enzyme or a proprotein sequence. Such fragments, derivatives and analogs are deemed to be within the scope of those skilled in the art from the teachings herein.

The enzymes and polynucleotides of the present invention are preferably provided in an isolated form, and preferably are purified to homogeneity.

The term "isolated" means that the material is removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide or enzyme present in a living animal is not isolated, but the same polynucleotide or enzyme, separated from some or all of the coexisting materials in the natural system, is isolated. Such polynucleotides could be part of a vector and/or such polynucleotides or enzymes could be part of a composition, and still be isolated in that such vector or composition is not part of its natural environment.

The enzymes of the present invention include the enzymes of SEQ ID NOS: 15-28 and 61-64 (in particular the mature enzyme) as well as enzymes which have at least 70% similarity (preferably at least 70% identity) to the enzymes of SEQ ID NOS: 15-28 and 61-64 and more preferably at least 90% similarity (more preferably at least 90% identity) to the enzymes of SEQ ID NOS: 15-28 and 61-64 and still more preferably at least 95% similarity (still more preferably at least 95% identity) to the enzymes of SEQ ID NOS: 15-28 and 61-64 and also include portions of such enzymes with such portion of the enzyme generally containing at least 30 amino acids and more preferably at least 50 amino acids.

As known in the art "similarity" between two enzymes is determined by comparing the amino acid sequence and its conserved amino acid substitutes of one enzyme to the sequence of a second enzyme.

A variant, i.e. a "fragment", "analog" or "derivative" polypeptide, and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions, fusions and truncations, which may be present in any combination.

Among preferred variants are those that vary from a reference by conservative amino acid substitutions. Such substitutions are those that substitute a given amino acid in a polypeptide by another amino acid of like characteristics. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu and Ile; interchange of the hydroxyl residues Ser and Thr, exchange of the acidic residues Asp and Glu, substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg and replacements among the aromatic residues Phe, Tyr.

15 Most highly preferred are variants which retain the same biological function and activity as the reference polypeptide from which it varies.

Fragments or portions of the enzymes of the present invention may be employed for producing the corresponding full-length enzyme by peptide synthesis; therefore, the fragments may be employed as intermediates for producing the full-length enzymes.

20 Fragments or portions of the polynucleotides of the present invention may be used to synthesize full-length polynucleotides of the present invention.

The present invention also relates to vectors which include polynucleotides of the present invention, host cells which are genetically engineered with vectors of the invention and the production of enzymes of the invention by recombinant techniques.

Host cells are genetically engineered (transduced or transformed or transfected) with the vectors of this invention which may be, for example, a cloning vector or an expression vector. The vector may be, for example, in the form of a plasmid, a viral particle, a phage, etc. The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants or amplifying the genes of the present invention. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to the ordinarily skilled artisan.

The polynucleotides of the present invention may be employed for producing enzymes by recombinant techniques. Thus, for example, the polynucleotide may be included in any one of a variety of expression vectors for expressing an enzyme. Such vectors include chromosomal, nonchromosomal and synthetic DNA sequences, e.g., derivatives of SV40; bacterial plasmids; phage DNA; baculovirus; yeast plasmids; vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia, adenovirus, fowl pox virus, and pseudorabies. However, any other vector may be used as long as it is replicable and viable in the host.

The appropriate DNA sequence may be inserted into the vector by a variety of procedures. In general, the DNA sequence is inserted into an appropriate restriction endonuclease site(s) by procedures known in the art. Such procedures and others are deemed to be within the scope of those skilled in the art.

The DNA sequence in the expression vector is operatively linked to an appropriate expression control sequence(s) (promoter) to direct mRNA synthesis. As representative examples of such promoters, there may be mentioned: LTR or SV40 promoter, the <u>E. coli. lac</u> or <u>trp</u>, the phage lambda P_L promoter and other promoters known to control expression of genes in prokaryotic or eukaryotic cells or their viruses. The expression

vector also contains a ribosome binding site for translation initiation and a transcription terminator. The vector may also include appropriate sequences for amplifying expression.

In addition, the expression vectors preferably contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells such as dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, or such as tetracycline or ampicillin resistance in <u>E. coli</u>.

The vector containing the appropriate DNA sequence as hereinabove described, as well as an appropriate promoter or control sequence, may be employed to transform an appropriate host to permit the host to express the protein.

As representative examples of appropriate hosts, there may be mentioned: bacterial cells, such as <u>E. coli</u>, <u>Streptomyces</u>, <u>Bacillus subtilis</u>; fungal cells, such as yeast; insect cells such as <u>Drosophila S2</u> and <u>Spodoptera Sf9</u>; animal cells such as CHO, COS or Bowes melanoma; adenoviruses; plant cells, etc. The selection of an appropriate host is deemed to be within the scope of those skilled in the art from the teachings herein.

More particularly, the present invention also includes recombinant constructs comprising one or more of the sequences as broadly described above. The constructs comprise a vector, such as a plasmid or viral vector, into which a sequence of the invention has been inserted, in a forward or reverse orientation. In a preferred aspect of this embodiment, the construct further comprises regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available. The following vectors are provided by way of example; Bacterial: pQE70, pQE60, pQE-9 (Qiagen), pD10, psiX174, pBluescript II KS, pNH8A, pNH16a, pNH18A, pNH46A (Stratagene); ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia); Eukaryotic: pSV2CAT, pOG44,

pXT1, pSG (Stratagene) pSVK3, pBPV, pMSG, pSVL (Pharmacia). However, any other plasmid or vector may be used as long as they are replicable and viable in the host.

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda P_R, P_L and trp. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

In a further embodiment, the present invention relates to host cells containing the above-described constructs. The host cell can be a higher eukaryotic cell, such as a mammalian cell, or a lower eukaryotic cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-Dextran mediated transfection, or electroporation (Davis, L., Dibner, M., Battey, I., Basic Methods in Molecular Biology, (1986)).

The constructs in host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence. Alternatively, the enzymes of the invention can be synthetically produced by conventional peptide synthesizers.

- 20 Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook, et al., Molecular Cloning: A Laboratory
- 25 Manual, Second Edition, Cold Spring Harbor, N.Y., (1989), the disclosure of which is hereby incorporated by reference.

Transcription of the DNA encoding the enzymes of the present invention by higher eukaryotes is increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp that act on a promoter to increase its transcription. Examples include the SV40 enhancer on the late side of the replication origin bp 100 to 270, a cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers.

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, e.g., the ampicillin resistance gene of E. coli and S. cerevisiae TRP1 gene, and a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), α-factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated enzyme. Optionally, the heterologous sequence can encode a fusion enzyme including an N-terminal identification peptide imparting desired characteristics, e.g., stabilization or simplified purification of expressed recombinant product.

Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host. Suitable prokaryotic hosts for transformation include <u>E. coli</u>, <u>Bacillus subtilis</u>, <u>Salmonella typhimurium</u> and various species within the genera Pseudomonas, Streptomyces, and Staphylococcus, although others may also be employed as a matter of choice.

As a representative but nonlimiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM1 (Promega Biotec, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced by appropriate means (e.g., temperature shift or chemical induction) and cells are cultured for an additional period.

Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents, such methods are well known to those skilled in the art.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, Cell, 23:175 (1981), and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

15

The enzyme can be recovered and purified from recombinant cell cultures by methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps.

The enzymes of the present invention may be a naturally purified product, or a product of chemical synthetic procedures, or produced by recombinant techniques from a prokaryotic or eukaryotic host (for example, by bacterial, yeast, higher plant, insect and mammalian cells in culture). Depending upon the host employed in a recombinant production procedure, the enzymes of the present invention may be glycosylated or may be non-glycosylated. Enzymes of the invention may or may not also include an initial methionine amino acid residue.

β-galactosidase hydrolyzes lactose to galactose and glucose. Accordingly, the OC1/4V, 9N2-31B/G, AEDII12RA-18B/G and F1-12G enzymes may be employed in the food processing industry for the production of low lactose content milk and for the production of galactose or glucose from lactose contained in whey obtained in a large amount as a by-product in the production of cheese. Generally, it is desired that enzymes used in food processing, such as the aforementioned β-galactosidases, be stable at elevated temperatures to help prevent microbial contamination.

These enzymes may also be employed in the pharmaceutical industry. The enzymes are used to treat intolerance to lactose. In this case, a thermostable enzyme is desired, as well. Thermostable β -galactosidases also have uses in diagnostic applications, where they are employed as reporter molecules.

Glucosidases act on soluble cellooligosaccharides from the non-reducing end to give glucose as the sole product. Glucanases (endo- and exo-) act in the depolymerization of cellulose, generating more non-reducing ends (endo-glucanases, for instance, act on internal linkages yielding cellobiose, glucose and cellooligosaccharides as products). β- glucosidases are used in applications where glucose is the desired product. Accordingly, M11TL, F1-12G, GC74-22G, MSB8-6G, OC1/4V, VC1-7G1, 9N2-31B/G and AEDII12RA18B/G may be employed in a wide variety of industrial applications, including in corn wet milling for the separation of starch and gluten, in the fruit industry for clarification and equipment maintenance, in baking for viscosity reduction, in the textile industry for the processing of blue jeans, and in the detergent industry as an additive. For these and other applications, thermostable enzymes are desirable.

Antibodies generated against the enzymes corresponding to a sequence of the present invention can be obtained by direct injection of the enzymes into an animal or by administering the enzymes to an animal, preferably a nonhuman. The antibody so obtained will then bind the enzymes itself. In this manner, even a sequence encoding only a fragment of the enzymes can be used to generate antibodies binding the whole native enzymes. Such antibodies can then be used to isolate the enzyme from cells expressing that enzyme.

For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler and Milstein, 1975, Nature, 256:495-497), the trioma technique, the human B-cell hybridoma technique (Kozbor et al., 1983, Immunology Today 4:72), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole, et al., 1985, in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96).

Techniques described for the production of single chain antibodies (U.S. Patent 4,946,778) can be adapted to produce single chain antibodies to immunogenic enzyme products of this invention. Also, transgenic mice may be used to express humanized antibodies to immunogenic enzyme products of this invention.

- Antibodies generated against the enzyme of the present invention may be used in screening for similar enzymes from other organisms and samples. Such screening techniques are known in the art, for example, one such screening assay is described in "Methods for Measuring Cellulase Activities", *Methods in enzymology*, Vol 160, pp. 87-116, which is hereby incorporated by reference in its entirety.
- The present invention will be further described with reference to the following examples; however, it is to be understood that the present invention is not limited to such examples. All parts or amounts, unless otherwise specified, are by weight.
 - In order to facilitate understanding of the following examples certain frequently occurring methods and/or terms will be described.
- "Plasmids" are designated by a lower case p preceded and/or followed by capital letters and/or numbers. The starting plasmids herein are either commercially available, publicly available on an unrestricted basis, or can be constructed from available plasmids in accord with published procedures. In addition, equivalent plasmids to those described are known in the art and will be apparent to the ordinarily skilled artisan.
- 20 "Digestion" of DNA refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The various restriction enzymes used herein are commercially available and their reaction conditions, cofactors and other requirements were used as would be known to the ordinarily skilled artisan. For analytical purposes, typically 1 µg of plasmid or DNA fragment is used with about 2 units of enzyme in about 20 µl of buffer solution. For the purpose of isolating DNA

fragments for plasmid construction, typically 5 to 50 µg of DNA are digested with 20 to 250 units of enzyme in a larger volume. Appropriate buffers and substrate amounts for particular restriction enzymes are specified by the manufacturer. Incubation times of about 1 hour at 37°C are ordinarily used, but may vary in accordance with the supplier's instructions. After digestion the reaction is electrophoresed directly on a polyacrylamide gel to isolate the desired fragment.

Size separation of the cleaved fragments is performed using 8 percent polyacrylamide gel described by Goeddel, D. et al., Nucleic Acids Res., 8:4057 (1980).

"Oligonucleotides" refers to either a single stranded polydeoxynucleotide or two complementary polydeoxynucleotide strands which may be chemically synthesized. Such synthetic oligonucleotides have no 5' phosphate and thus will not ligate to another oligonucleotide without adding a phosphate with an ATP in the presence of a kinase. A synthetic oligonucleotide will ligate to a fragment that has not been dephosphorylated.

"Ligation" refers to the process of forming phosphodiester bonds between two double stranded nucleic acid fragments (Maniatis, T., et al., Id., p. 146). Unless otherwise provided, ligation may be accomplished using known buffers and conditions with 10 units of T4 DNA ligase ("ligase") per 0.5 µg of approximately equimolar amounts of the DNA fragments to be ligated.

Unless otherwise stated, transformation was performed as described in the method of 20 Graham, F. and Van der Eb, A., Virology, 52:456-457 (1973).

Example 1

Bacterial Expression and Purification of Glycosidase Enzymes

DNA encoding the enzymes of the present invention, SEQ ID NOS: 1-14 and 57-60 were initially amplified from a pBluescript vector containing the DNA by the PCR technique using the primers noted herein. The amplified sequences were then inserted into the

respective PQE vector listed beneath the primer sequences, and the enzyme was expressed according to the protocols set forth herein. The 5' and 3' primer sequences for the respective genes are as follows:

Thermococcus AEDII12RA -18B/G

5 5' CCGAGAATTCATTAAAGAGGAGAAATTAACTATGGTGAATGCTATGATTGTC 3' (SEQ ID NO:29)

3' CGGAAGATCTTCATAGCTCCGGAAGCCCATA 5' (SEQ ID NO:30)

Vector: pQE12; and contains the following restriction enzyme sites 5' EcoRI and 3' Blg II.

OC1/4V-33B/G

- 5' CCGAGAATTCATTAAAGAGGAGAAATTAACTATGATAAGAAGGTCCGATTTTCC 3' (SEQ ID NO:31)
 - 3' CGGAAGATCTTTAAGATTTTAGAAATTCCTT 5' (SEQ ID NO:32)

Vector: pQE12; and contains the following restriction enzyme sites 5' EcoRI and 3' Bgl Π.

- 15 Thermococcus 9N2 31B/G
 - 5' CCGAGAATTCATTAAAGAGGAGAAATTAACTATGCTACCAGAAGGCTTTCTC 3' (SEQ ID NO:33)
 - 3' CGGAGGTACCTCACCCAAGTCCGAACTTCTC 5' (SEQ ID NO.34)

Vector: pQE30; and contains the following restriction enzyme sites 5' EcoRI and 3'

20 KpnI.

Staphylothermus marinus F1 - 12G

- 5' CCGAGAATTCATTAAAGAGGAGAAATTAACTATGATAAGGTTTCCTGATTAT 3' (SEQ ID NO:35)
- 3' CGGAAGATCTTTATTCGAGGTTCTTTAATCC 5' (SEQ ID NO:36)
- Vector: pQE12; and contains the following restriction enzyme sites 5' EcoRI and 3' Bgl II.

Thermococcus chitonophagus GC74 - 22G
5' CCGAGAATTCATTAAAGAGGAGAAATTAACTATGCTTCCAGGAGAACTTTCTC 3'
(SEQ ID NO:37)

3' CGGAGGATCCCTACCCCTCTTAAGATCTC 5' (SEQ ID NO:38)

Vector: pQE12; and contains the following restriction enzyme sites 5' EcoRI and 3' BamHI.

M11TL

5' AATAATCTAGAGCATGCAATTCCCCAAAGACTTCATGATAG 3' (SEQ ID NO:39)
3' AATAAAAGCTTACTGGATCAGTGTAAGATGCT 5' (SEQ ID NO:40)
Vector: pQE70; and contains the following restriction enzyme sites 5' Sphl and 3'
Hind III.

Thermotoga maritima MSB8-6G

5' CCGACAATTGATTAAAGAGGAGAAATTAACTATGGAAAGGATCGATGAAATT 3' (SEQ ID NO:41)
3' CGGAGGTACCTCATGGTTTGAATCTCTTCTC 5' (SEQ ID NO:42)

Vector: pQE12; and contains the following restriction enzyme sites 5' EcoRI and 3' KpnI.

Pyrococcus furiosus VC1 - 7G1

5' CCGACAATTGATTAAAGAGGAGAAATTAACTATGTTCCCTGAAAAGTTCCTT 3' (SEQ ID NO:43)
3' CGGAGGTACCTCATCCCCTCAGCAATTCCTC 5' (SEQ ID NO:44)
Vector: pQE12; and contains the following restriction enzyme sites 5' EcoRI and 3'
Kpn I.

Bankia gouldi endoglucanase (37GP1)

5' AATAAGGATCCGTTTAGCGACGCTCGC 3' (SEQ ID NO:45)
3' AATAAAAGCTTCCGGGTTGTACAGCGGTAATAGGC 5' (SEQ ID NO:46)
Vector: pQE52; and contains the following restriction enzyme sites 5' Bam HI and 3' Hind III.

Thermotoga maritima α-galactosidase (6GC2)

5' TTTATTGAATTCATTAAAGAGGAGAAATTAACTATGATCTGTGTGGAAATATTCGGAAAG 3' (SEQ ID NO:47)

3' TCTATAAAGCTTTCATTCTCTCACCCTCTTCGTAGAAG 5' (SEQ ID NO:48)

5 Vector: pQET; and contains the following restriction enzyme sites 5' EcoRI and 3' Hind III.

Thermotoga maritima \(\beta\)-mannanase (6GP2)

5' TTTATTCAATTGATTAAAGAGGAGAAATTAACTATGGGGATTGGTGGCGACGAC 3' (SEQ ID NO:49)

10 3' TTTATTAAGCTTATCTTTTCATATTCACATACCTCC 5' (SEQ ID NO:50)

Vector: pQEt; and contains the following restriction enzyme sites 5' Hind III and 3' EcoRI.

AEPII 1a β-mannanase (63GB1)

5' TTTATTGAATTCATTAAAGAGGAGAAATTAACTATGCTACCAGAAGAGTTCCTATGGGGC 3'

15 (SEQ ID NO:51)

3' TTTATTAAGCTTCTCATCAACGGCTATGGTCTTCATTTC 5' (SEQ ID NO:52)

Vector: pQEt; and contains the following restriction enzyme sites 5' Hind III and 3' EcoRI.

OC1/4V endoglucanase (33GP1)

20 5'

AAAAAACAATTGAATTCATTAAAGAGGAGAAATTAACTATGGTAGAAAGACACTTCAGATATGTTCT T 3' (SEQ ID NO:53)

3' TTTTTCGGATCCAATTCTTCATTTACTCTTTGCCTG 5' (SEQ ID NO:54)

Vector: pQEt; and contains the following restriction enzyme sites 5' BamHI and 3'

25 EcoRI.

Thermotoga maritima pullalanase (6GP3)

5' TTTTGGAATTCATTAAAGAGGAGAAATTAACTATGGAACTGATCATAGAAGGTTAC 3' (SEQ ID NO:55)

3' ATAAGAAGCTTTTCACTCTCTGTACAGAACGTACGC 5' (SEQ ID NO:56)

Vector: pQEt; and contains the following restriction enzyme sites 5' EcoRI and 3' Hind III.

The restriction enzyme sites indicated correspond to the restriction enzyme sites on the bacterial expression vector indicated for the respective gene (Qiagen, Inc. Chatsworth, CA). The pQE vector encodes antibiotic resistance (Amp'), a bacterial origin of replication (ori), an IPTG-regulatable promoter operator (P/O), a ribosome binding site (RBS), a 6-His tag and restriction enzyme sites.

The pQE vector was digested with the restriction enzymes indicated. The amplified sequences were ligated into the respective pQE vector and inserted in frame with the sequence encoding for the RBS. The ligation mixture was then used to transform the E. coli strain M15/pREP4 (Qiagen, Inc.) by electroporation. M15/pREP4 contains multiple copies of the plasmid pREP4, which expresses the lacI repressor and also confers kanamycin resistance (Kan'). Transformants were identified by their ability to grow on LB plates and ampicillin/kanamycin resistant colonies were selected. Plasmid DNA was isolated and confirmed by restriction analysis. Clones containing the desired constructs were grown overnight (O/N) in liquid culture in LB media supplemented with both Amp (100 ug/ml) and Kan (25 ug/ml). The O/N culture was used to inoculate a large culture at a ratio of 1:100 to 1:250. The cells were grown to an optical density 600 (O.D. 600) of between 0.4 and 0.6. IPTG ("Isopropyl-B-D-thiogalacto pyranoside") was then added to a final concentration of 1 mM. IPTG induces by inactivating the lacI repressor, clearing the P/O leading to increased gene expression. Cells were grown an extra 3 to 4 hours. Cells were then harvested by centrifugation.

The primer sequences set out above may also be employed to isolate the target gene from the deposited material by hybridization techniques described above.

Example 2

Isolation of A Selected Clone From the Deposited genomic clones

A clone is isolated directly by screening the deposited material using the oligonucleotide primers set forth in Example 1 for the particular gene desired to be isolated. The specific oligonucleotides are synthesized using an Applied Biosystems DNA synthesizer. The oligonucleotides are labeled with ³²P- -ATP using T4 polynucleotide kinase and purified according to a standard protocol (Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, Cold Spring, NY, 1982). The deposited clones in the pBluescript vectors may be employed to

transform bacterial hosts which are then plated on 1.5% agar plates to the density of 20,000-50,000 pfu/150 mm plate. These plates are screened using Nylon membranes according to the standard screening protocol (Stratagene, 1993). Specifically, the Nylon membrane with denatured and fixed DNA is prehybridized in 6 x SSC, 20 mM NaH₂PO₄, 0.4%SDS, 5 x Denhardt's 500 μg/ml denatured, sonicated salmon sperm

DNA; and 6 x SSC, 0.1% SDS. After one hour of prehybridization, the membrane is hybridized with hybridization buffer 6xSSC, 20 mM NaH₂PO₄, 0.4%SDS, 500 ug/ml denatured, sonicated salmon sperm DNA with 1x10⁶ cpm/ml ³²P-probe overnight at 42°C. The membrane is washed at 45-50°C with washing buffer 6 x SSC, 0.1% SDS for 20-30 minutes dried and exposed to Kodak X-ray film overnight. Positive clones

are isolated and purified by secondary and tertiary screening. The purified clone is sequenced to verify its identity to the primer sequence.

Once the clone is isolated, the two oligonucleotide primers corresponding to the gene of interest are used to amplify the gene from the deposited material. A polymerase chain reaction is carried out in 25 µl of reaction mixture with 0.5 ug of the DNA of the gene of interest. The reaction mixture is 1.5-5 mM MgCl₂, 0.01% (w/v) gelatin, 20 µM each of dATP, dCTP, dGTP, dTTP, 25 pmol of each primer and 0.25 Unit of Taq polymerase. Thirty five cycles of PCR (denaturation at 94°C for 1 min; annealing at 55°C for 1 min; elongation at 72°C for 1 min) are performed with the Perkin-Elmer Cetus automated thermal cycler. The amplified product is analyzed by

agarose gel electrophoresis and the DNA band with expected molecular weight is excised and purified. The PCR product is verified to be the gene of interest by subcloning and sequencing the DNA product. The ends of the newly purified genes are nucleotide sequenced to identify full length sequences. Complete sequencing of full length genes is then performed by Exonuclease III digestion or primer walking.

Example 3

Screening for Galactosidase Activity

Screening procedures for α -galactosidase protein activity may be assayed for as follows:

Substrate plates were provided by a standard plating procedure. Dilute XL1-Blue MRF *E coli* host of (Stratagene Cloning Systems, La Jolla, CA) to O.D.₆₀₀ = 1.0 with NZY media. In 15 ml tubes, inoculate 200 μl diluted host cells with phage. Mix gently and incubate tubes at 37 °C for 15 min. Add approximately 3.5 ml LB top agarose (0.7%) containing 1mM IPTG to each tube and pour onto all NYZ plate surface. Allow to cool and incubate at 37 °C overnight. The assay plates are obtained as substrate p-Nitrophenyl α-galactosidase (Sigma) (200 mg/100 ml) (100 mM NaCl, 100 mM Potassium-Phosphate) 1% (w/v) agarose. The plaques are overlayed with nitrocellulose and incubated at 4 °C for 30 minutes whereupon the nitrocellulose is removed and overlayed onto the substrate plates. The substrate

Example 4

Screening of Clones for Mannanase Activity

A solid phase screening assay was utilized as a primary screening method to test clones for \(\beta \)-mannanase activity.

25 A culture solution of the Y1090-E. coli host strain (Stratagene Cloning Systems, La Jolla, CA) was diluted to O.D.₆₀₀=1.0 with NZY media. The amplified library from Thermotoga maritima lambda gtl1 library was diluted in SM (phage dilution buffer):

 5×10^7 pfw/µl diluted 1:1000 then 1:100 to 5×10^2 pfw/µl. Then 8 µl of phage dilution (5×10^2 pfw/µl) was plated in 200 µl host cells. They were then incubated in 15 ml tubes at 37 °C for 15 minutes.

Approximately 4 ml of molten, LB top agarose (0.7%) at approximately 52 °C was added to each tube and the mixture was poured onto the surface of LB agar plates. The agar plates were then incubated at 37 °C for five hours. The plates were replicated and induced with 10 mM IPTG-soaked Duralon-UVTM nylon membranes (Stratagene Cloning Systems, La Jolla, CA) overnight. The nylon membranes and plates were marked with a needle to keep their orientation and the nylon membranes were then removed and stored at 4 °C.

An Azo-galactomannan overlay was applied to the LB plates containing the lambda plaques. The overlay contains 1% agarose, 50 mM potassium-phosphate buffer pH 7, 0.4% Azocarob-galactomannan. (Megazyme, Australia). The plates were incubated at 72 °C. The Azocarob-galactomannan treated plates were observed after 4 hours then returned to incubation overnight. Putative positives were identified by clearing zones on the Azocarob-galactomannan plates. Two positive clones were observed.

The nylon membranes referred to above, which correspond to the positive clones were retrieved, oriented over the plate and the portions matching the locations of the clearing zones for positive clones were cut out. Phage was eluted from the membrane cut-out portions by soaking the individual portions in 500 µl SM (phage dilution buffer) and 25 µl CHCl₃.

Example 5

Screening of Clones for Mannosidase Activity

A solid phase screening assay was utilized as a primary screening method to test clones for β-mannosidase activity.

A culture solution of the Y1090-E. coli host strain (Stratagene Cloning Systems, La Jolla, CA) was diluted to O.D.600=1.0 with NZY media. The amplified library from AEPII 1a lambda gtl1 library was diluted in SM (phage dilution buffer): 5 x 10⁷ pfu/ μ l diluted 1:1000 then 1:100 to 5 x 10² pfu/ μ l. Then 8 μ l of phage dilution 5 (5 x 10^2 pfu/µl) was plated in 200 µl host cells. They were then incubated in 15 ml tubes at 37 °C for 15 minutes.

Approximately 4 ml of molten, LB top agarose (0.7%) at approximately 52 °C was added to each tube and the mixture was poured onto the surface of LB agar plates. The agar plates were then incubated at 37 °C for five hours. The plates were replicated and induced with 10 mM IPTG-soaked Duralon-UV™ nylon membranes (Stratagene Cloning Systems, La Jolla, CA) overnight. The nylon membranes and plates were marked with a needle to keep their orientation and the nylon membranes were then removed and stored at 4 °C.

A p-nitrophenyl-\(\beta\)-D-manno-pyranoside overlay was applied to the LB plates containing the lambda plaques. The overlay contains 1% agarose, 50 mM potassiumphosphate buffer pH 7, 0.4% p-nitrophenyl-ß-D-manno-pyranoside. (Megazyme, Australia). The plates were incubated at 72 °C. The p-nitrophenyl-ß-D-mannopyranoside treated plates were observed after 4 hours then returned to incubation overnight. Putative positives were identified by clearing zones on the p-nitrophenyl-

β-D-manno-pyranoside plates. Two positive clones were observed. 20

The nylon membranes referred to above, which correspond to the positive clones were retrieved, oriented over the plate and the portions matching the locations of the clearing zones for positive clones wre cut out. Phage was eluted from the membrane cut-out portions by soaking the individual portions in 500 µl SM (phage dilution buffer) and 25 µl CHCl₃.

Example 6

Screening for Pullulanase Activity

Screening procedures for pullulanase protein activity may be assayed for as follows:

Substrate plates were provided by a standard plating procedure. Host cells are diluted to $O.D._{600} = 1.0$ with NZY or appropriate media. In 15 ml tubes, inoculate 200 μ l diluted host cells with phage. Mix gently and incubate tubes at 37 °C for 15 min. Add approximately 3.5 ml LB top agarose (0.7%) is added to each tube and the mixture is plated, allowed to cool, and incubated at 37 °C for about 28 hours.

Overlays of 4.5 mls of the following substrate are poured:

10	100 ml to	tal volume
	0.5g	Red Pullulan Red (Megazyme, Australia)
	1.0g	Agarose
	5ml	Buffer (Tris-HCL pH 7.2 @ 75 °C)
	2ml	5M NaCl
15	5ml	CaCl ₂ (100mM)
	85ml	dH₂O

Plates are cooled at room temperature, and thenm incubated at 75°C for 2 hours. Positives are observed as showing substrate degradation.

Example 7

20

Screening for Endoglucanase Activity

Screening procedures for endoglucanase protein activity may be assayed for as follows:

- The gene library is plated onto 6 LB/GelRite/0.1% CMC/NZY agar plates
 (~4,800 plaque forming units/plate) in E.coli host with LB agarose as top agarose.
 The plates are incubated at 37°C overnight.
 - 2. Plates are chilled at 4°C for one hour.
 - 3. The plates are overlayed with Duralon membranes (Stratagene) at

room temperature for one hour and the membranes are oriented and lifted off the plates and stored at 4°C.

- 4. The top agarose layer is removed and plates are incubated at 37°C for ~3 hours.
- 5. The plate surface is rinsed with NaCl.
 - 6. The plate is stained with 0.1% Congo Red for 15 minutes.
 - 7. The plate is destained with 1M NaCl.
- 8. The putative positives identified on plate are isolated from the Duralon membrane (positives are identified by clearing zones around clones). The
 10 phage is eluted from the membrane by incubating in 500μl SM + 25μl CHCl₃ to elute.
 - 9. Insert DNA is subcloned into any appropriate cloning vector and subclones are reassayed for CMCase activity using the following protocol:
 - i) Spin 1ml overnight miniprep of clone at maximum speed for 3 minutes.
- ii) Decant the supernatant and use it to fill "wells" that have been made in an LB/GelRite/0.1% CMC plate.
 - iii) Incubate at 37°C for 2 hours.
 - iv) Stain with 0.1% Congo Red for 15 minutes.
 - v) Destain with 1M NaCl for 15 minutes.
 - vi) Identify positives by clearing zone around clone.

Numerous modifications and variations of the present invention are possible in light of the above teachings and, therefore, within the scope of the appended claims, the invention may be practiced otherwise than as particularly described.

20

WHAT IS CLAIMED IS:

- 1. An isolated polynucleotide selected from the group consisting of:
 - (a) SEQ ID NOS: 1-14 and 57-60;
 - (b) SEQ ID NOS: 1-14 and 57-60, wherein T can also be U;
 - (c) polynucleotide sequences complementary to SEQ ID NOS: 1-14 and 57- 60;
 - (d) polynucleotide sequences which encode an amino acid sequence as set forth in SEQ ID NOS:15-28, and 61-64; and
 - (e) fragments of (a), (b), (c) or (d) that are at least 15 consecutive bases in length and that will selectively hybridize to DNA which encodes a polypeptide of SEQ ID NOS:15-28, and 61-64.
- 2. A vector comprising a polynucleotide of claim 1.
- 3. A host cell containing the vector of claim 2.
- 4. The method of claim 3, wherein the host cell is a eukaryotic cell.
- 5. The method of claim 3, wherein the host cell is a prokaryotic cell.
- 6. A method for producing a polypeptide comprising:
 - (a) culturing the host cells of claim 3;
 - (b) expressing from the host cell of claim 3 a polypeptide encoded by said polynucleotide; and
 - (c) isolating the polypeptide.

7. An enzyme selected from the group consisting of:

- (a) an enzyme comprising an amino acid sequence set forth in SEQ ID
 NOS: 15-28 or 61-64; and
- (b) an enzyme which comprises at least 30 consecutive amino acid residue as an enzyme of (a).
- 8. An enzyme of which at least a portion is coded for by a polynucleotide of claim 1, and which is selected from the group consisting of:
 - (a) an enzyme comprising an amino acid sequence which is at least 70% identical to an amino acid sequence selected from the group of amino acid sequences set forth in SEQ ID NOS:15-28 or 61-64; and
 - (b) an enzyme which comprises at least 30 amino acid residues to the enzyme of (a).
- 9. A method for generating glucose from soluble cell oligosaccharides comprising contacting a sample containing oligosaccharides with an effective amount of an enyzme selected from the group consisting of an enzyme having the amino acid sequence set forth in SEQ ID NOS: 15-28, 61-63 and 64 such that glucose is produced.
- 10. The method of cliam 9, wherein the sample is selected from the group consisting of dairy products, fruit juices, detergents, textiles, guar gum, animal feed, plant biomass and waste products.
- 11. The method of claim 9, wherein the oligosaccharide is selected from the group consisting of maltose, cellobiose, lactose, sucrose, raffinose, stachyose, verbascose, cellulose, starch, amylose, glycogen, disacharrides, polysacharrides and pullulan.

M11TL GLYCOSIDASE - 29G COMPLETE GENE SEQUENCE - 9/95

	1 17	T: A	A 10														٠.							
	1 M	4 13	//: 15 //: 15	ha t		^^ (i	VI. J.	J.f. V	,TT.	ATA	(2)	. TA	. 44.	.v .	٠ -١- ١	11.Y	177	; 77	-r (*/	10 7	-J-I.	٠:٨	ለ ፡፡፡ ግ	4.13
					• • • • • • • • • • • • • • • • • • • •	y 2+ A	sp r	114. P	•••	11	1217	· TY		1 :	•••	; ;	1,1	, m		100 10	1	1:1	A (27) 11 Alii	
11	1 12	רא ינה	m e	a. G	3: T (ar es	A(: (:	AT 1	T TE : .	A A T	ALT	" I:A	T 74.	· •									: CAC	
:	1 61	y 11	e P	rn G	ly Se	er G	lu A	sp P	1	Asn	See	Ass	ייד מ	n 7			717	: 67	1; (7	VT G	V.J.	.,,	C CAC	1.20
12		K: A(,	:y (:(:A (3	.T C	: A C	ra c	1.L. V	ın. e	33.	1:AT	. 17-	r can	e c	N;)	\AC	care	. cr	A (::	T	M.,	77.7	: ۸۸1	ino
. 4	l As	n Ti	13 V	I A	la G	ly L	eu V	al s	er (:1 y	Ang	Phi	e Pr	o C	lu A	na/	Gly	Pr	o G1	v T	V 1	T.,	: AAT 1 AHD	100
181																								
61	Le	^ ~	n C1	\ \ \ \ \	VT CA	ים טי	AC EL	אכי בי	TC: C	:: T	CVC	AAC	C.10	CCC	C C	TT	AAC	. VC	ተ ለኅ	T A	iλ	CTA	v GGC	240
•	•	" "		A:	in As	р н	15 A:	sp l.	eu /	(1a	Clu	Lys	Le	n C1	y v	'a l	Asn	Th	r 11	e A	rg	va l	GCC	80
241																								
8 1	. ,Va	1 61	u Tr	p Se	r Ar	o 1)	- Di		1	VIC	B	The	TTO		T C	77	***	CT	: cc	TC	ra i	CVC	AGA Arg	300
						• • •		ie r		.y >	L. 1. 0	in	Pho	e as	n V	a l	Lys	Va.	l Pr	o Va	11	Slu	Arg	100
301	GA	T GA	GAA	c cc	C AG	C AT	7 67	T C	AC C	TA	GAT	CTC	· CAT	r ca	T .			~					GAT	,
101	λs	p Gl	u As	n Gl	y Se	r Il	e Va	1 H	is V	al	λsp	Val	Ass	. un	ה ה	~~	415	Uni	CA	A AC	iA (777	SAT ASD	360
																								120
361	GA.	A TT	A GC	C YY	C AA	G GA	G GC	C C	TA A	λC	CAT	TAC	CTA	(GA	A A	TG	TAT	AA.	GA				~ .	420
121	C1	Le	u Al	a As	n Ly	s C1	u Al	a Va	ıl A	sn.	His	Tyr	Val	G1	u M	et	Tyr	Lys	Ası	o Tr	70 V	/n }	C)	420 140
423																								140
421		٠ (زن:	r AG	A AA	A CI	T AT	y CI	CN	TT	TA '	TAC	CAT	TGG	CC	c c.	TG	CCT	CTC	TG	CT	T	'nC	AAC	480
	- A+ 1	, 01	, ~1	g Ly	s Le	u II	e Le	u As	תו	en ,	Tyr	His	Trp	Pr	o L	eu	Pro	Leu	Tr) Le	u F	is	Asn	160
481																								
161	Pro	Ile	. Me	t Va	G AG	7 2 2	~ Ma	5 GU			200	AGA	GCG	CC	C T(CA (CCC.	TCC	CTY	KA 1	CG	AG	GAG	540
					,.	y	y ne	. 01	y r		nsp	viâ	V10	PT	0 50	er (GIY	Trp	Lev) As	ם מ	lu	Clu	180
541	TCC	GTC	CT	CA	3 77	r GC	C AA	A TA	c 6	cc (GCA	TAC	ATT	·	T TY	202		100						
181	Sei	Va]	Va.	1 G1:	Ph:	e Ala	a Ly	s Ty	T A	la d	Ala	Tyr	Ile	Ala	. T:	~ .	Lve	Mer	Cla			- A	CCT	600 200
																								200
601	CII	ATC	TC	S YC	ACC	2 AT	G AA	د ه	A C	ככ א	AC	CTC	CTT	TAT	C.D	\G (CAA	GGA	TAC	. AT	G T	TC	GTT	660
201	Val	Het	TI	Se	Thi	r Mei	LAS	n Gl	u P	ro A	\sn	۷al	Val	Туз	: G1	u c	iln	Gly	Tyr	Me	t P	he	Val	220
661																								
221	7.45	Cla	C1.	TTC	CC	, cc	: GCI	TA	כ די	K S	CT	TTC	Gλλ	CCI	, ec	T	λŢ	AAG	CCC	AG	G A	Çλ	AAT	720
	Lys	01 ,	U 13	Pne	Pro	PIC	. C1	Y TY	rL	eu 5	Ser	Leu	Glu	λla	r YJ	.a. 7	SP	Lys	Ala	Ar	7	rg	Asn	240
721	ATG	ATC	CAC		CAT	י פרו			~ #1															
241	Met	Ile	Glr	λla	His	ιλla	Arc	, GC		· ·		~~ 1	71-				TC	AGT	AAG	w	, C		CII	780
				-				, ,,,_,	,	•			116	Lys	, AI	y r	116	SET	Lys	Lyı	P	.0	Val	260
781	CGA	CTA	ATA	TAC	CCT	. TTC	: cx	TC	3 17	c G	λλ	CTA	TTA	GAG	GG	T C	CA	CCA	GAA	CT)			ChT	840
261	Gly	Leu	Ile	Tyr	Ala	Phe	Clr	TI	Ph	e G	lu	Leu	Leu	Glu	Gl	y P	TO	Ala	Glu	Va)	P)		Asn	280
841	AAG	111	XXQ	AGC	TCT	AAC	TT	TAC	TA	TT	TC .	ACA	GAC	ATA	GT.	A T	CG .	AAG	GCT	AGT	· T(:λ	ATC	900
281	Lys	Phe	Lys	5er	Ser	Lys	Leu	Ty	עז	r P	he '	Thr	Yżb	Ile	Va.	1 S	er i	Lys	Gly	Ser	Se	r	lle	300
901																								
301	Ile	Asn	Val	Clin	TAC	AGG	AGA	GAT	CT	T G	CC /	AAT	AGG	CTA	GA	בי	GG :	TTC	GGC	CII	N	c:	TAC	960
-					Tyr	. ALY	AI 9	vzt	, re	. .	14 /	ASn	Arg	Leu	AS	P T	י סדי	Leu	Gly	Val	As	n	lyr	320
961	TAT	AGC	CGT	TTA	CTC	TAC	AAA	ATC	GT	c G	AT (GAC		دس	271		TC /	-T-C	C > C					1020
321	Tyr	Ser	Arg	Leu	Va]	Tyr	Lys	Ile	۷۵	1 A:	5D /	\sp	Lvs	Pro	Tle	- T	ie i		ui.	Cly	17		-1	340
																								340
1021	TTC	CIT	TCT	ACA	CCI	CCC	CCC	ATC	AG	c c	C (CT (GAA .	AAT	CCT	r TO	T)	\GC	GAT	TIT	GC	c :	rcc	1080
341	Phe	Leu	Cys	Thr	Pro	Cla	Gly	Ile	Se	r Pi	ro j	la	Glu .	Asn	Pro	C	ys S	Ser .	Asp	Phe	C1	y 1	rp	360
												•												
361	GAG	U-1	TAT	CCT	CVY	GGA	CTC	TAC	CT.	A C	ידי כ	TA .	AAA I	GAA	CTT	T	/C Y	uc :	CGA	TAC	GC	GC	TA	1140
201	210	√a;	ryt	PTO	G) u	CIA	Leu	Tyr	Le	ı Le	eu L	eu i	Lys	Clu	ren	T)	/F A	Sn.	Arg	туr	C1	y V	/a]	380
141	GAC	Tro	V.Lc.	CTY:	A(''	(:A/·		~~~	,															
381	GAC Asp	Leu	He	Val	The	6)	Acc	Ch	Uni	1 TC		AC /	NGC A	ACK:	GAT	CC	G 7	TC	AGA	CCC	CC	λT	,YC	1200
	. – •-				Thr	.,,,,,,	A>0	217	va i	. 50	: . ^	sp :	er i	n rq	\SI	, v;	a L	eu .	Arg	Pro	A1	a 7	,À I	400
201	CTC	GTC	TCG	CAT	ידיהו	3. VC.	ACC	ATD	TCI	. AA	M C	cc z	:/ =r -		1:40	,,,	٠, ٠							
401	Leu	Val.	Ser	nis	Val	Tyi	Ser	Val	Tre	L	2. J	la	120	A :: 11	(:1	1:1		15 (····	GTC	,^^	л G	GC T	1260
																								420
261	TAC	CTt.	LVI.	rea:	ve:c.	7776:	VCV	CAL	۲۸۸	. TA	c c	A(; 1	na: c	яч.	r'Ati	cic	.c. 1.	rı.	ver:	(·A··	A A	A 7-	-17 ·	13.20
421	Tyr	lanı	Hi:	Tiji	See	lanı	This	Asp	۱۰۰	ту	r 1:	10.7	111	١.,	Clin	GI	y /·	1100	ro :	Glai	1.0	. 1	lu.	440

Figure la

1441	Th.	(YE:	CAC:			:		•			. ,		V1.1	1,4		A 1	1.000	1771. V.13	4
1441	CAG	TAA	14	46			•	•	•••	 VMI)		1,11	Gln		147,	.Thr	Levis	110	4110

Figure 1b(Continued)

OC1/4 GLYCOSIDASE - 33G/B COMPLETE GENE SEQUENCE - 9/95

	I ATK ATA ACA ACA
	Het Ile Arn Aud TCC GAT TTT (Y'A AAA GAT TTT ATC
	I ATC ATA AGA AGG TCC GAT TTT CCA AAA GAT TTT ATC TTC GGA ACG GCT ACG GCA GCA TAC 60 61 CAG ATT GAA GGT GCA GCA AAC CAA GAT GAT GAT GCA ACG GCT ACG GCA ACG GCA TAC 60 61 CAG ATT GAA GGT GCA GCA AAC CAA GAT GAT GAT GAT GAT GAT GAT GAT GAT G
	"I CAG ATT CAA COM
	Gin Tie Giu Gly Ala Ala Asn Giu Asp Giy Arg Giy Pro Ser Tie Trp Asp Val Phe Ser 40
•	Ala Asn Glu Asp Gly Arg Gly Pro Ser 11 TGG CAT GTC TTT TCA 120
	TAL CAC ACC CCT con
	41 His The Pro Gly Lys The Act GCT GAC ACA GGA GAC GTT GCC
	121 CAC ACG CCT GGC AAA ACC CTG AAC GGT GAC ACA GGA GAC GTT GCG TCT GAC CAT TAT CAC 180 181 CGA TAC AAG GAA GAT ATC CAC CTC LEG ACC GTT GLY AEP VAL ALA CYS ASP HIS TYF HIS 60
	THE COA TAC AND ON THE EA
	61 Arg Tyr Lys Glu Asp Ile Gln Leu Met Lys Glu Ile Gly Leu Asp Ala Tyr Arg Phe Ser 80
	241 Leu Het Lys Glu Ile Gly Leu Asp Ala Tyr Act TCT 240
	ATC TCC TGG CCC AGA ATT ATG CCA COM COM
	241 ATC TCC TGG CCC AGA ATT ATG CCA GAT GGG AAG AAC ATC AAC CAA AAG GGT GTG GAT TTC 300 301 TAC AAC AGA CTC GTT GAT GAG GTT GAT GAG GTT GAT GAG GAT GA
	301 TAC AND ACT OF STATE OF ST
	301 TAC AAC AGA CTC GTT GAT GAG CTT TTG AAG AAT GAT ATC ATA CCA TTC GTA ACA CTC TAT 360
	101 TYP ASR APR LEW Val ASP Glu Lew Lew Lys Asr APP Ile Ile Pro Phe Val Thr Lew Typ 120
	361 CAC TOO GOO THE TAT 360
	161 CAC TGG GAC TTA CCC TAC GCA CTT TAT GAA AAA GGT GGA TGG CTT AAC CCA GAT ATA GCG 420
	121 His Trp Asp Leu Pro Tyr Ala Leu Tyr Glu Lys Gly Gly Trp Leu Asn Pro Asp Ile Ala 140
	421 CTC TAT THE ASP ILE ALA 140
	421 CTC TAT TTC AGA GCA TAC GCA ACG TTT ATG TTC AAC GAA CTC GGT GAT CGT GTG AAA CAT 480 481 TGG ATT ACA CTG AAC GAA GCA TGC TGC AAC GAA CTC GGT GAT AGG GAA CAT 480 481 TGG ATT ACA CTG AAC GAA GGA TGCA TGCA T
	Ale Tyr Ale Thr Phe Het Phe Ash Clu Leu Clu Art CGT GTG ANA CAT 480
	481 TGG ATT ACT THE VALLEYS HIS 160
	481 TGG ATT ACA CTG AAC GAA CCA TGG TGT TCT TCT TTC TCG GGT TAT TAC ACG GGA GAG CAT 540
	161 Trp Ile Thr Leu Ash Glu Pro Trp Cys Ser Ser Phe Ser Gly Tyr Tyr Thr Gly Glu His 180
	541 GCC CCG GGT CAT CAA AAT TTA CAA GAA GCG ATA ATC GCG GCG CAC AAC CTG TTG AGG GAA . 600
	Ala Pro Gly His Gln Asn Law Gla GCG ATA ATC GCG GCG CAC AAC CTG TTG ACC GA
	181 Ala Pro Gly His Gln Asn Leu Gln Glu Ala Ile Ile Ala Ala His Asn Leu Leu Arg Glu 200 601 CAT GGA CAT GCC GTC CAG GGG GGA GGA ARG GGG GGA ARG GGG GAA
	601 CAT GGA CAT GCC GTC CAG GCG TCC AGA GAA GAA GAA GAT GGG GAA GTT GGC TTA ACC 660 201 His Gly His Ala Val Gln Ala Ser Arg Glu Glu Val Lys Asp Gly Glu Val Co
	201 His Gly His Ala Val Gln Ala Ser Arg Glu Glu Val Lys Asp Gly Glu Val Gly Leu Thr 220 661 AAC GTT GTG ATG AAA ATA GLA GAT GGC TTA ACC 220
	661 AAC CTT COO 100 VAI GLY Leu Thr 220
	661 AAC GTT GTG ATG AAA ATA GAA CCG GGC GAT GCA AAA CCC GAA AGT TTC TTG GTC GCA AGT 720 221 ASN Val Val Het Lys Ile Glu Pro Gly Asp Ala Lys Pro Glu Ser Phe Leu Val Ala Ser 240
	THE LYS ILE GIU Pro Gly Asp Ala Lys Pro Gly Ser TO TTG GTC GCA AGT 720
	721 CIT CIT CIT CIT CIT IN 12 240
	721 CTT GTT GAT AAG TTC GTT AAT GCA TGG TCC CAT GAC CCT GTT GTT TTC GGA AAA TAT CCC 780 781 GAA GAA GCA GTT GCA CTT TTC SCA CTT GTT GTA TAT CCC 780
	As Ala Trp Ser His Asp Pro Val. Val. Phe Cly Line 780
	781 GAA GAA GCA GTT GCA CTT TAT ACG GAA AAA GGG TTG CAA GTT CTC GAT AGG GAT ATG AAT 840
	261 Glu Glu Ala Val Ala Leu Tyr Thr Glu Lys Gly Leu Gln Val Leu Asp Ser Asp Het Asn 280 841 ATT ATT TCG ACT CCT ATA GROUP AND GROUP AND GROUP ASP SER ASP HET ASN 280
	841 ATT ATT TOG ACT CO. 280
	841 ATT ATT TCG ACT CCT ATA GAC TTC TTT GGT GTG AAT TAT TAC ACA AGA ACA CTT GTT GTT 900
	281 Ile Ile Ser Thr Pro Ile Asp Phe Phe Gly Val Asn Tyr Tyr Thr Arg Thr Leu Val Val 300
	901 TTT CAT amp and 100
. :	901 TIT GAT ATG AAC AAT CCT CIT GGA TIT TCG TAT GIT CAG GGA GAC CIT CCC AAA ACG GAG 960 961 ATG GGA TGG GAA ATG TAG GGG TAT GTG Val Gln Gly Asp Leu Pro Lys Thr Glu 320
	Ash FFO Leu Gly Phe Ser Tyr Val Gln Gly Ash Lou GAA ACG GAG 960
5	961 ATG GGA TOC CALL AND 120
٠.	Het Gly Trp Glu lle Tyr Pro Gln Gly Leu Phe Asp Het Leu Val Tyr Leu Lys Glu Arg 340
10	1020 The Gin Gly Leu Phe Asp Met Leu Val Tyr Leu Lux Clu
.,	21 TAT AAA CTA CCA CTT TAT ATC ACA GAG AAC GGG ATG GCT GGA CCT GAT AAA TTG GAA AAC 1080
	TYPE LYS Leu Pro Leu Tyr Ile Thr Glu Asn Gly Met Ala Cly Day AAA TTG GAA AAC 1080
10	41 Tyr Lys Leu Pro Leu Tyr Ile Thr Glu Asn Gly Het Ala Gly Pro Asp Lys Leu Glu Asn 360
3	61 GIV ATT CAT GAT AAT TAC CGA ATT GAA TAT TOO COL
	81 GGA AGA GTT CAT GAT AAT TAC CGA ATT GAA TAT TTG GAA AAG CAC TTT GAA AAA GCA CTT 1140 41 GAA GCA ATC AAT GCA GAT GTT GAT GAT GAT GAT GAT GAT GAT GA
114	41 GAA CCA ATC AND THE SID LYS AIR LEU 380
3 8	41 GAA GCA ATC AAT GCA GAT GTT GAT TTG AAA GGT TAC TTC ATT TGG TCT TTG ATG GAT AAC 1200
	All Asp Val Asp Leu Lys Cly Tyr Phe Ile Try Car ATG GAT AAC 1200
120	TTC GAA TCC CIC TOO 400
40	Phe Glu Trp Ala Cys Gly Tyr San Tro GGT ATA ATC TAC GTA CAT TIC SAN CAT
126	Phe Glu Trp Ala Cys Gly Tyr Ser Lys arg Phe Gly Ile Ile Tyr Val Asp Tyr Asn Thr 420
42	1 CCA ANA AGG ATA TIG ANA GAT TOA COC AND DECEMBER AND THE 420
72	1 CCA AAA AGG ATA TTG AAA GAT TCA CCC ATC TCC TTG AAG GAA TTT CTA AAA TCT TAA 1317 Pro Lys Arg 11c lau Lys Asp Ser Ala Her Trp Leu Lys Glu Blook 1
	Pro Lys Arg He Leu Lys Asp Ser Ala Met Trp Leu Lys Glu Phe Leu Lys Ser End 419
	· · · · · · · · · · · · · · · · · · ·

Figure 2

STAPHYLOTHERMUS MARINUS GLYCOSIDASE -COMPLETE GENE SEQUENCE 9/95

1 TTC AMA AGE	
1 TTG ATA AGG TIT CCT GAT TAT TTC TRU TIT GOA ACA GGT AGA TCA TCG GAC CAG ATY. 1 Het lie Arg Phe Pro Asp Tyr Phe Leu Phe Gly Thr Ala Thr Ser See Nor Phe	
1 Het Ile Arg Phe Pro Asp Tyr Phe Leu Phe Gly Thr Ala Thr Ser Ser His Gln Ile	GA: 60
61 CTT AND THE GIR LIE	Glu 20
61 GGT AAT AAC ATA TIT AAT GAT TGG TGG GAG TGG GAG ACT AAA GGC AGG ATT AAG GTG 21 Gly Asn Asn Ile Phe Asn Asp Trp Trp Glu Trp Glu Thr Lys Gly Arg Ile Asn GTG	
AST AST THE PHE AST ASP TEP TEP CITY TO CALL AND GCC ACG ATT AND GTG.	ACA 120
121 TCG CGT AAC CO	Ara 40
121 TCG GGT AAG GCA TGT AAT CAT TGG GAA CTC TAT AAA GAA GAC ATA GAG CTT ATG GGT G 181 CTG GGA TAT AAT GCT TAT AGG TGG GAA CTC TAT AAA GAA GAC ATA GAG CTT ATG GGT G 181 CTG GGA TAT AAT GCT TAT AGG TGG	
Ala Cys Asn His Trp Glu Leu Tyr Lys Glu Asn GAG CTT ATG GCT C	CAC 180
181 CTG GGA TAT AND COM	ilu 60
61 Leu Gly Tyr Asn Ala Tyr Arg Phe Ser Ile Glu Trp Ser Arg Ile Phe Pro Arg Lys A	
Ala lyr Arg Phe Ser Ile Glu Trp Ser Arg Ile Phi CCC AGA AMA G	AT 240
241 CAT ATA CAT THE CAT	5D 80
81 His Ile ASP TYI GIV SAT LAG TAT AAG GAA ATA GTT AAT CTA	•
81 His Ile ASP TYP Glu Ser Leu ASD Lys Tyr Lys Glu Ile Val ASD Leu AND Lys Tyr Lys Glu Ile Val ASD Lys Tyr Lys Tyr Lys Glu Ile Val ASD Lys Tyr Lys	AC 300
301 GGG ATA GAA CCT GTA ATC ACT CTT CAC CAC TTC ACA AAC CCG CAA TGG TTT ATG AAA AT	YF 100
101 Gly 11e Glu Pro Val 11e Thr Low CAC CAC TTC ACA AAC CCG CAA TGG TTT ATC ALL	
101 Gly Ile Glu Pro Val Ile Thr Leu His His Phe Thr Asn Pro Gln Trp Phe Het Lys Il	360
JOI GGT GGA TGG ACT AGG GAA GAG AAC ATTA ALL THE AND THE BYE IT	le 120
361 GGT GGA TGG ACT AGG GAA GAG AAC ATA AAA TAT TTT ATA AAA TAT GTA GAA CTT ATA GG 121 Gly Gly Trp Thr Arg Glu Glu Asn Ile Lys Tyr Phe Ile Lys Tyr Val Glu Leu Ile Al 421 TCC GAG ATA AAA GAC CTC LLA AND AND AND AND AND AND AND AND AND AN	T 420
421 TOO GOOD TO THE LYS TYP Val Glu Leu Ile Al	a 140
421 TCC GAG ATA AAA GAC GTG AAA ATA TGG ATC ACT ATT AAT GAA CCA ATA ATA TAT GTT TT. 141 Ser Glu Ile Lys Asp Val Lys Ile TTP Ile Thr Ile Asp Glu Pro Ile TTP TTP TTP TTP TTP TTP TTP TTP TTP TT	40
141 Ser Glu Ile Lys Asp Val Lys Ile Trp Ile Thr Ile Asn Glu Pro Ile Ile Tyr Val Les 481 CAA GGA TAT ATT TCC GGG GLA TATA	A 480
481 CAA GCD THE ACT TO THE ACT OF THE TENT OF THE TENT VAL Let	160
481 CAA GGA TAT ATT TCC GGC GAA TGG CCA CCT GGA ATT AAA AAT TTA AAA ATA GCT GAT CAI 161 Gln Gly Tyr Ile Ser Gly Glu Trp Pro Pro Gly Ile Lys Asn Leu Lye Tla Aaa CAT GAT CAI	
161 Gln Gly Tyr Ile Ser Gly Glu Trp Pro Pro Gly Ile Lys Asn Leu Lys Ile Aia Asp Glr 541 GTA ACT ANG ANT CTT TTO AND ACT TO AND ACT TO AND ACT TO AND ACT TO	A 540
541 GTA ACT AIG AND G	180
181 Val The Lys Asn Leu Leu Lys Ala His Asn Glu Ala Tyr Asn Ile Leu His Lys His Gly 601 ATT GTA GGC ATA GCT Lab and the control of the contro	
Deu Lys Ala His Asn Glu Ala Tyr Asn Ile Leu Min His Line	600
601 ATT GTA GGC ATA GCT ANA ANC ATG ATA GCA TIT ANA CCA GGA TCT ANT AGA GGA ANA GAC 201 Ile Val Gly Ile Ala Lys Asn Het Ile Ala Phe Lys Pro Gly Ser Asn And GAC	200
201 Ile Val Gly Ile Ala Lys Asn Het Ile Ala Phe Lys Pro Gly Ser Asn Arg Gly Lys Asp 661 ATT AAT ATT TAT CAT LIA COR CUE	
the Ala Phe Lys Pro Gly Ser Asn Arg Gly Lys Asn	660
661 ATT AAT ATT TAT CAT ANA GTC GAT AAA GCA TTC AAC TGG GGA TTT CTC AAC GGA ATA TTA 221 Ile Asn Ile Tyr His Lys Val Asp Lys Ala Phe Asn Trp Gly Phe Levy Art Atta	220
THE AER HE TYT HIS LYS VAL ASD LYS ALE PHO ARE TOG GGA TIT CTC AAC GGA ATA TTA	720
221 Ile Asn Ile Tyr His Lys Val Asp Lys Ala Phe Asn Trp Gly Phe Leu Asn Gly Ile Leu 721 AGG GGA GAA CTA GAA ATT TO	240
721 AGG GGA GAA CTA GAA ACT CTC CGT GGA AAA TAC CGA GTT GAG CCC GGA AAT ATT GAT TTC 241 Arg Gly Glu Leu Glu Thr Leu Arg Gly Lys Tyr Arg Val Glu Pro Gly Arg TX	
ory of the Leu Glu Thr Leu Arg Gly Lys Tyr Arg Val Glu Ben Gly Art ATT GAT TTC	780
781 ATA GGC ATA AAC TIE TO THE ASP Phe	260
261 Ile Gly Ile Asn Tyr Tyr Ser Ser Tyr Ile Val Lys Tyr Thr Trp Asn Pro Phe Lys Leu 841 CAT ATT AAA GTC GAA COL TO THE TYP ASN PRO Phe Lys Leu	
THE ASP PER SET TYPE I LEVEL LYS TYPE THE TEN ASP PER THE ASP PER THE ASP PER THE THE ASP PER THE THE ASP PER THE THE TEN ASP PER THE	840
841 CAT ATT AAA GTC GAA CCA TTA GAT ACA GGT CTA TGG ACA ACT ATG GGT TAC TGC ATA TAT 281 His Ile Lys Val Glu Pro Leu Asp Thr Gly Leu Trp Thr Thr Her Gly Dec ATA TAT	280
281 His Ile Lys Val Glu Pro Leu Asp Thr Gly Leu Trp Thr Thr Het Gly Tyr Cys Ile Tyr 901 CCT AGA GGA ATA TAT CAL COM	000
901 CCT AGA CGT and The Gly Leu Trp Thr Thr Het Gly Tyr Cys Ile Tyr	900
901 CCT AGA GGA ATA TAT GAA GTT GTA ATG AAA ACT CAT GAG AAA TAC GGC AAA GAA ATA ATC 301 Pro Arg Gly Ile Tyr Glu Val Val Het Lys Thr His Glu Lys Tyr Gly Lys Tyr Gly Lys	300
JOI Pro Arg Gly Ile Tyr Glu Val Val Het Lys Thr His Glu Lys Tyr Gly Lys Glu Ile Ile	960
961 ATT ACA CAG SAG	320
961 ATT ACA GAG AAC GGT GTT GGA GTA GAA AAT GAT GAA TTA AGG ATT TTA TCC ATT ATC AGG 121 Ile Thr Glu Asn Gly Val Ala Val Glu Asn Asp Glu Leu Arg Ile Leu Son The Company of	
J21 Ile Thr Glu Asn Gly Val Ala Val Glu Asn Asp Glu Leu Arg Ile Leu Ser Ile Ile Arg	1020
1021 CAC TTA CAR TIC TIE TIE Arg	340
1021 CAC TTA CAA TAC TTA TAT AAA GCC ATG AAT GAA GGA GCA AAG GTG AAA GGA TAT TTC TAC 141 His Leu Gln Tyr Leu Tyr Lys Ala Het Asn Glu Gly Ala Lys Val Lys Gly Tyr Phe Tyr 1081 TGG AGC TTC ATG GAT AND THE	
THE LYS Ala Het Asn Glu Cly Ala Lys Val Lys Cly And THE THE	1080
1081 TGG AGC TTC ATC GAT AAT TTT GAG TGG GAT AAA GGA TTT AAC CAA AGG TTC GGA CTA GTA 161 Trp Ser Phe Hec Asp Asn Phe Glu Trp Asp Lys Gly Phe Asn Gln Arg Phe Glu	360
161 Trp Ser Phe Het Asp Asn Phe Glu Trp Asp Lys Gly Phe Asn Gln Arg Phe Gly Leu Val	1140
THE SID FTP ASP LYS Cly Phe Asn Cln Arg Phe Cly Lin Val	1140 380
381 Glu Val ASP TYR LYS The Phe Glu Arg Lys Pro Arg Lys Ser Ala Tyr Val Tyr Ser Gln 1201 ATA GCA CGT ACC ANG AST	1200
1201 ATA CICA COT LOS	400
401 LIE ALL COT ACC ANG ACT ATA ACT GAT GAN TAC CTA CAN AND THE	
1201 ATA GCA CGT ACC AAG ACT ATA AGT GAT GAA TAC CTA GAA AAA TAT GGA TTA AAG AAC CTC 401 Ile Ala Arg Thr Lys Thr Ile Ser Asp Glu Tyr Leu Glu Lys Tyr Glu	1260
1261 CAA TO LEU LYS ASD Leu	420
421 Glu End 422	•
·	

Figure 3

Thermococcus 9N2 Glycosidase -318/G Complete gene sequence 9/95

										-		• • • •		-10-11	,	1 73							
		;		. (+)		N GAU	, ecc	TITE :	. 610	TO	G (G)	ירם ב	TC		~							ccc	
		•		LAL	PEC	o Cli	, CJ)	/ Phe	Leu	TEE	01	/ Val	Ser	Cla			111	CAG	TTC	GAC	ATC	CCC	60
		1	٠							• •					. 241	Ely	Ph•	CIU	Phe	Glu	Net	CCC	20
	2	- 1	1	****	כזכ	. AC	, voc	: AAC	A:T	CAT	י כנינ		ACA	CAC	-							•	
	-	•	~•p	ryı	Leu	YES	, yrd	A A ETT	Ile	A.P) F. c	A.sen	The	Am	***	700	we	TCC	CLC	ACC	CAT	. ccc	120
	121																						40
	41																						
	•	•		ABD	Ile	Lys	γīβ	Cln	≟eu	Val	Ser	Glv	300		cee	CYC	CYC	CCC	ATA	AAC	iic	TAC	180
	101																						60
	401	` `	•••	Cit	TAC	CYC	AAC	GAT	CAC	CGC	CTC	C:C:C	31.0	~		cor							•
	61		10	Leu	IXI	Cla	Lys	Asp	HAR	725	Lev	7.7	~~~	CAC	CIC	eja egi	CLC	γyc	CII	TAC	λŒ	ATT	240
																							80
	241	C	CA	ATA	CYC	ce	ACC	YCC	ATC	Ť						TIT				_	-		. 80
	, BJ	0	JA	Ile.	Cla	TIP	Sez	YE a	II.	Dha		- 00	-	YCE	7.GG	TIT	CTC .	cvc	CTT	GAC	مستدي	C) C	• • • •
																							300
																							100
	101	λ	rg .	λεp	Ser	علائ	Gly	Lou	Val	Tue	Mar.	GIC	7 57	ATC	Cat	YYY (GAC :	ACG (czć i	GAA	CAC		
																							360
	361																						120
	12:)	\$ D (31u	Ile.	Ala.	λza	H			ATA	CCC	TAC	TAC	CCC	CGC (3 TT 3	LTA (: מגד	-24-	~~		
•																							420
	421																						140
	141	C	lu I	44	Cly .	Pho :	Lys	Val :	71- 1	-16	.	נינט :	MC (CAC :	TTC :	ACC C	c	CC 0	T .	· .	-		
																							480
																							160
	161	25	PP	=0 ;	1.	lie i	11.	c	~~()	uc (307	ג אד:	icc x	uc c	CT)	ra I	TT C	GC T	CC C	~ ~			
																							540
																							180
	181	G1	u S	er V	al v	701 0	:\ r	TO C	ice y	AG 1	CAC (3000	CC 1	אכ א	TC G	77 Y 50 Y	æ	- 4-	TY ::	~ ~			
																							600
																							200
:	201	Va.	ىد 1	א ספ	er 7	~ ·	3C A	T	TC A	YC 3	ive c	<u> </u>	<u> </u>	3C C	II G	TG CO	AG (**	ب م					
																							660
																							220
2	21	PE	- T	/Y 5	er c	lv n	rr c	<u> </u>	CC C	3C C	II Y	TG A	1 C C	CC 6	4 5 G	בה פר דעיק		· -	<u> </u>				
																							720
7	21	NAC.		Y: 1	T2 2	. ~ ~											J	- 4	14 A.	a I	le L	5.7	240
2	41	λs:	Me	+ T		AL 150	טע ט	42 GC	:X C7	<i>7</i> C C	C T	rc N	G AT	נג כז	נג גז	על פּי		~					
																							780
7	81	œ	- ==	T 1.2											•	てお		- ~=	b va	a 'va	u ty	*	260
2	52 .	ملة	À	5 · ·	~ U		- C C C	ic ro	C GY	9 00	بد ه	VG 67	C C3	K AT	יג גי		-						
																							840
																							280
2	31 ;	Ala	7			~ 54		ב א	د هم	C CC	עעני	ح حد	c 67	c x	x 60	- CC:	8 Ca:				_		
				- · -	y	* ~3	D 24	I Am	22 A.S.	D LI	O Ly	3 AE	P Va	l Ly	# A1	a 37:	. 63.			- XX	כ בא	c 9	00
9 (1 1	TC	Cad	·										•						P A.C.	2 LA	r	100
30	1 1	he.	p: .					متت ع	د هم	= 6 C	A AT	C CA	C AA	G GC	C AA	cro	- 330						
•	-			. 36	. 64	y _a	u Ph	e Ph	رمير د	P Al.	11	e Hi	Lvs	s C1	v Lva	CIC		- ALI	CAC	3 1.1.0	C CX	C 9	60
96	1 0	CT	CAC														. ~-:	111	976) Pbe	A A S	> 3	20
32	1 6	1~	61.	71	- 77T	- GT		1 (2)		CA:	r cm	CAC	000	AAC	C3/	TEP							
	_			LA	PIM	· VA	Ly	Va)	LArg	ni:	Lei	, ,,,,	CIV	/ AET		~~~	, MIM		: מזי	. YYC	TA	1	020
102	3 ~	20	.~			_						,	,				114	GTA	. A93	Ası	1 TY:	. 3	40 .
-,-	- •		1.11	virg	615	1 VA	L Vaj	Arg	TYT	Sez	- 631	Pro	Lva	Phe	Pro	500	7.7	ccc	CIG	ATA	TC	: 1	080
108	1												-, •		••••	AGC Sex	TTE	PIO	Leu	Ila	Sei	: 3	60
36	•	• •	_~	الانتاب																			
	- ~ .		neg.	CIA	VAI	Hie	Arn	TYE	CJA	IVI	A:	Çyĸ	Ara	PEA	Gla	ACT Sej	TOT	TCC	CCC	CYC	CCY	. 1:	140
3141												•			,	344	SEL	Ser	AIA	1	~ >		90
383		~ ,		GTA	ACC		2			_													
	~ ~ =	. ų 1	FIO	Val	5cr	Asp	Tle	Gly	TIP	612	11+	Tyr	Pro	Cla	Cliv	ATC Ile	IAC	באב	TCC	YIY	YCY	12	200
1201	6 2								_			• •			0.y	*14	LAE	ALD.	Ser	Ile	۾ ته	40	0
401				AAC.	AAA	TAC	~~~	~															
		ر ت	. 7 9	A-271	Lys	224	Cly	Val	Pro	Va 1	TYE	۷al	The	61"	7.40	ĊJA CCY	ATA	ccc	CXT	TCA	XCT	12	60
1261											•			014	AL 13	CIY	110	A : A	7	C	~~-	42	0
421	_	_ ^		CTG	CCC	~-	***																
1	~	ד ע	T.F	-41	Arg	Pro	TYT	Tyr	Leu	Ala	Ser	Kie	Val	Al-	7.22	ATT Ile		CYC	CCE	TAC	CYC	1)	20
					•									~ •	⊷y∎.	TTB	Clu	Clu	Ala	Tyr	Olu	44	0

Figure 4a

1321 441 -381	CCC	: cc																			
441	Ala	Gli		C CA	CIC	, YCC	C-20	20	٠							٠.					
		O1,	′ ту	, yai) Val	λro	GI	. ~~		TAC	LCC	CCC	: כזנ	ACC	CAC						
.381	~~~					_	,	.71	Len	: 1.\1	طني	Ala	Leu	The	>	~~	TAC	: cu	100	CCC	1380
401	Leu	Cly	Phe	Are	Her		1.16	COC	CIC	TAT	AAA	CTG	CAT						-		***
461 1441 481						7.0	Phe	CIA	ربهنا	Tys	Lve	Val		CIC	ATA	YCC	AAC	CAG	AG.	ALA	
																					1440
481	Pro	Arm	C)		AGC	CTA	AAC	CTT	TAT	3.00	000							-20	₩.	Thr	480
			O1U	Glu	Ser	Val	Lya	741	~~~	-60		ATC	CTC	CYC	AAC	AAC	CCN				
150:	C 1 1						-,-		. 7 .	viå	G:A	I.u	Val	Glu	Asp	À ===	~~~	OIC	YOC	W C	1500
																	CIA	Val	5er	Lys	500
501	CIU	Ile.	Arg	Clu	Lve	Pho	~~	CIT	CCC	TOA	15	30								-	550
					-,-	, nd	GIY	- 40	Glv	Rnd		_									

Figure 4b(Continued)

20

120

ATG GAA AGG ATC GAT GAA ATT CTC TCT CAG TTA ACT ACA GAG GAA AAG GTG AAG Giu lle Leu Ser Gin CTC CIT Leu Thr The Civ Ċш 1.74 1.cu Val GTG GGG GIT GGT CTT CCA GGA CTT TTT GGG AAC CCA CAT TCC AGA GTG 61 Val Gly Leu CCG CCT Gly Leu Phe CCC CCT Gly Asn Pro His Ser X1A Val ۸ta GIV Ala GGA GAA ACA CAT CCC GTT CCA AGA CTT GGA ATT CCT GCG 40 121 TIT כדכ כדם Gly Glu The His CCA Pro Val CAT CCT Pro Arg Leu. Cly He Pro 180 Ala Λb Asp Gly Pro ഹ GCA GGA CTC AGA ATA AAT CCC ACA AGG GAA AAC GAT GAA AAC ACT TAC Ala Gly Leu Arg He Asn Pro Thr Arg Glu TAC ACG ACG GCA 240 Asp Asp Ciu Thr Thr Thr TIT CCC GTT GAA ATC ATG CTC GCT TCT ACC TGG AAC AGA GAC CTT Ala CTG Val Glu lic CAA GAA Met CTG Leu Ala Ser ` Thr GGA 300 Τmp Ass Arg Aso Leu Giv Clu V۵I Gly 100 AAA GCC ATG GGA GAA GAA GTT AGG GAA TAC GGT GTC GAT GTG CTT 301 101 CTT Mci Gly Glu GCA Val Arg Giu Tyr CCT GCG ATG 360 Gly Val A.sp Val Ala Pro Met ANC ATT CAC AGA AAC CCT CTT TGT GGA AGG AAT TTC GAG TAC TAC TCA 120 361 121 His Arg Asn GAA GAT CCT Leu Cys Gly Arg Asn Phe CTC 420 Glu . Tyr Tyr Ser Giu Pro ٧al CIT TOO GOT GAN ATG GOT TOA GOO TIT GTO ANG GGA GIT CAN TOT 140 421 141 CAA GGG Mct Ala Scr Ala CTG GGA GCC 480 Phe Val Lys Gly Val Gìn Gln Gly Val Giy Ala TOC ATA AMA CAC TIT GTC GCG AAC AAC CAG GAA ACG AAC AGG ATG 160 481 161 CTA CTG Phe Vaj GAC AZA GZA ALA ACG ATC 540 Gla Glu The Asn Arg Met Val Val Asp Thr He GTG TCC GAG CGA GCC CTC AGA GAA ATA TAT CTG AAA GGT TTT 180 541 181 Glu Arg Ala GAA Leu ATT GCT CTC Arg Glu lie 600 Tyr Leu Lys Gly He Ala Val Lys Lys 200 GCA AGA CCC TGG ACC GTG ATG AGC GCT TAC AAC AAA CTG AAT GGA AAA 601 201 Val Met Ser TAC TOT TCA CAG 660 Ala Tyr Asn Lys Leu Gly Lys Cys Ser Gin 220 661 MC GM TGG CTT TTG ANG ANG CIT CTC AGG GAN GAN TGG GGA 221 Ash Glu III GGC Trp Leu CCT TTC Lys Val Leu ATG 720 Arg Glu Clu Ĩφ Gly Gly Mei AGE GAC TGG TAC GCG GGA GAE AAC CCT GTA GAA CAG CTC AAG GCC 240 721 241 Trp Tyr Gly Asp Asn AAC GAT ATG ATC 780 Pro Val Cin Leu Lys Gly Asn ATG CCT GGG AAA GCG TAT CAG GTG AAC ACA GAA AGA AGA GAT GAA ATA McI 781 261 Gin Val Asn GAA GAA ATC Thr Glu Arg Arg Giv Glu Hc 280 GAG GCG TTG AAG GAG GGA AAA TTG AGT GAG GAG GTT CTC GAT GAG TGT Gly Lys Leu Ser CTG AGA AAC ATT 900 Giu Leu Cys Val Arg CTC AAA GIT CTT GTG AAC GCG CCT TCC TTC AAA GGG TAC AGG TAC Asn 901 TCA Asn Ala Pro Ser AAC AAG CCG 960 Phe Lys Cly Tyr Arg Tyr Scr Asn Pro CTC GAA TCT CAC. GCG GAA GTC GCC TAC GAA GCA Lys 320 A₅p 961 321 CCT GCG GAG GGT Ser His GTI ۸b Clu Val Ala CTC CTT CTT GAG 1020 Tyr Ciu Gly Ala Clu Cly Val Val AAC AAC GGT GTT CTT CCG TTC GAT GAA AAT ACC CAT Lev Glu 1021 341 CTC Gly Val CCC टाट गां Leu Pro Phe Asp GCC CCT CAA 1080 Glu Axa The His Val ٧ы Pho Cly 360 ATC GAA ACA ATA AAG GGA GGA ACG GGA AGT GGA GAC 1081 Giy Cin 361 ACC CAT Thr lie CCG AGA 1.yx Gly Gly Thr TAC ACG ATC TCT 1140 Cly Ciy A.yı Tite Pru ATC CTT GAA GGC ATA AAA GAA AGA AAC ATG AAG ITC Tyr Thr lic Ser 380 Glu Gly He Lys Glu Arg Asn GAC GAA GAA CTC CCT TCC ACT TAT 1200 Mei Lys Phe Амр Clu Clu Ala Sei The Tyr 400

Figure:.5a

1												pr (-IV	Glu	Tyr	Lys	Pr		e l'i				rcc	1240
4:									,				uu j	ገገር	Lev	Ser	Ch	\A	A G	AG A	TA A	AG A	'ТР -^^ -уз	1320
								•		- •	•• ••	91 A	wi. I	ic	Ser	Arg	lic	C TC Ser	C)	•	•		AC yr	1380
							•		,	P '''	,	n u	.u 2	CT	Αsp	Aso	Clu	G CT			,		ÁA /S	1440
					-				- ~4		1 (1)	y L)	rs L	ys '	Val	Val	Val	1 en	1			rc G	5A	1500
						•					, ~	, 12	U V	N /	√ 20	Głv	He	CTT Leu	1					1560 520
							-	-			~~	~4	, v.	, ,	.cu	Val	Civ	Lys	P1	F AAT	CC Pro	•		1620 540
										-,-	بردت	1 77	20		33 0	Val	Pro	TCC Ser	TGC Trp		Phe			1680 560
											V 21	A #1	1 71	· G	łu	Ciu	A	11-	TAC Tyr	GTG Val	GG/ Giy	TA(1740 580
								•				~4	1 YT	G	. u	Pac	GGC Gly	TAC Tyr	GGC Gly	CTC Lev	TCT Ser	TAC Tyr	_	800 600
						•	_,_	. —		AAA Lys	110	A la	lle	۸.	D (Giv	Chu	ACG Thr	CTC Lev	AGA Arg	CTG Val	TCG Ser	-	860 620
								•		AGA Arg	714	Uiy	Lys	Gh	ı \	/a!	5	C1-	GTC Val	TAC	ATC Ne	AAA Lys		920 140
								•	-,-	Pro GAA Giu	FIRE	Cita	Cib	Lev	, i	VS A	4 la	DL-	CAC His	AAA Lys	ACA Thr	AAA Lys		80 60
2041	ACT	TTC										SET	Lev	Ciu	110	c F	,10	Lev	AGA Arg	GAT Asp	CTT ·	GCG Ala	20- 64	4 0
5101 5101 681	Ser				AT /	NTA ,	AGG ·	TG /	AGA		0.0	JCI	O/V	Glu	т.				AGG Arg	GTC Vai	GGT Gly	GCA Ala	210 70	
2161		TGA	216	6	1 da	le ,	Arg	Leu ,	Arg	Asp OA1	lic	Phe	Leu	Val	G)	NG G		-	AAG Lys	AGA Arg	TTC Phe	Lys	216 72	

Figure 5b(Continued)

THERMOCOCCUS AEDII12RA GLYCOSIDASE COMPLETE GENE BEQUENCE - 9/95 ATG ATC CAC TGC CCG GTT ANA GGG ATT ATA TCT GAG GCT CGC GGC ATA ACC ATC ACA ATA Met lie His Cys Pro Val Lys Gly lie lie Ser Glu Ala Arg Gly lie Thr lie Thr lie 60 61 GAT TTA AGT TIT CAA GGC CAA ATA AAT TTG GTG AAT GCT ATG ATT GTC TIT CCG GAG 20 Asp Leu Ser Phe Gin Cly Gin Ile Asn Asn Leu Val Asn Ala Het Ile Val Phe Pro Glu 120 121 TTC TTC CTC TTT GGA ACC GCC ACA TCT TCT CAT CAG ATC GAG GGA GAT AAT AAA TGG AAC 40 Phe Phe Leu Phe Gly Thr Ala Thr Ser Ser His Gln Ile Glu Gly Asp Asn Lys Trp Asn 180 CAC TGG TGG TAT TAT GAG GAG ATA GGT AAG CTC CCC TAC AAA TCC GGT AAA GCC TGC AAT 181 60 Asp Trp Trp Tyr Tyr Glu Glu Ile Gly Lys Leu Pro Tyr Lys Ser Gly Lys Ala Cys Asn 61 240 CAC TOG GAG CTT TAC AGG GAA GAT ATA GAG CTA ATG GCA CAG CTC GGC TAC AAT GCC TAC 80 His Trp Glu Leu Tyr Arg Clu Asp Ile Glu Leu Het Ala Gln Leu Gly Tyr Asn Ala Tyr 300 301 CGC TIT TCG ATA GAG TGG AGC CGT CTC TTC CCG GAA GAG GGC AAA TTC AAT GAA GAA GCC 100 101 Arg Phe Ser Ile Glu Trp Ser Arg Leu Phe Pro Glu Glu Gly Lys Phe Asn Glu Glu Ala TTC AAC CGC TAC CGT GAA ATA ATT GAA ATC CTC CTT GAG AAG GGG ATT ACT CCA AAC GTT 120 Phe Asn Arg Tyr Arg Glu Ile Ile Glu Ile Leu Leu Glu Lys Gly Ile Thr Pro Asn Val ACA CTG CAC CAC TTC ACA TCA CCG CTG TGG TTC ATG CGG AAG GGA GGC TTT TTG AAG GAA 140 Thr Leu His His Phe Thr Ser Pro Leu Trp Phe Het Arg Lys Gly Gly Phe Leu Lys Glu 480 CAA AAC CTC AAG TAC TGG GAG CAG TAC GTT GAT AAA GCC GCG GAG CTC CTC AAG GGA GTC 160 Glu Asn Leu Lys Tyr Trp Glu Gln Tyr Val Asp Lys Ala Ala Glu Leu Leu Lys Gly Val 540 ANG CIT GTA GCT ACA TTC ANC GAG CCG ATG GTC TAT GTT ATG ATG GGC TAC CTC ACA GCC 180 Lys Leu Val Ala Thr Phe Asn Glu Pro Het Val Tyr Val Het Het Gly Tyr Leu Thr Ala 600 דאכ זכה ככה ככה דוב אום אכן ככה דון אוא פכה דון אוא פון פכה פכא ואל כוב כין 200 Tyr Trp Pro Pro Phe Ile Lys Ser Pro Phe Lys Ala Phe Lys Val Ala Ala Asn Leu Leu 201 ANG GCC CAT GCA ATG GCA TAT GAT ATC CTC CAT GGT AAC TIT GAT GTG GGG ATA GTT AAA 220 661 Lys Ala His Ala Het Ala Tyr Asp Ile Leu His Gly Asn Phe Asp Val Gly Ile Val Lys 221 720 721 ARC ATC CCC ATA ATG CTC CCT GCA AGC ARC AGA GAG ARA GAC GTA GRA GCT GCC CAA ARG 240 Asn Ile Pro Ile Het Leu Pro Ala Ser Asn Arg Glu Lys Asp Val Glu Ala Ala Gln Lys 780 GCG GAT AAC CTC TIT AAC TGG AAC TTC CTT GAT GCA ATA TGG AGC GGA AAA TAT AAA GGA 260 781 Ala Asp Asn Leu Phe Asn Trp Asn Phe Leu Asp Ala Ile Trp Ser Gly Lys Tyr Lys Gly GCT TIT GGA ACT THE ANA ACT CON GAN AGO GAT GON GAC TTO ATH GGG ATH AND THE THE 841 Ala Phe Gly Thr Tyr Lys Thr Pro Glu Ser Asp Ala Asp Phe Ile Gly Ile Asn Tyr Tyr 281 900 300 ACA GCC AGC GAG GTA AGG CAT AGC TGG AAT CCG CTA AAG TIT TTC TTC GAT GCC AAG CTT Thr Ala Ser Glu Val Arg His Ser Trp Asn Pro Leu Lys Phe Phe Phe Asp Ala Lys Leu 960 GCA GAC TTA AGC GAG AGA AAA ACA GAT ATG GGT TGG AGT GTC TAT CCA AAG GGC ATA TAC 320 Ala Asp Leu Ser Glu Arg Lys Thr Asp Het Gly Trp Ser Val Tyr Pro Lys Gly Ile Tyr GAA GCT ATA GCA AAG GTT TCA CAC TAC GGA AAG CCA ATG TAC ATC ACG GAA AAC GGG ATA Glu Ala Ile Ala Lys Val Ser His Tyr Gly Lys Pro Het Tyr Ile Thr Glu Asn Gly Ile 1080 1081 GCT ACC TTA GAC GAT GAG TGG AGG ATA GAG TTT ATC ATC CAG CAC CTC CAG TAC GTT CAC 360 361 Ala Thr Leu Asp Asp Glu Trp Arg Ile Glu Phe Ile Ile Gln His Leu Gln Tyr Val His 1140 ANA GCC TTA ANC GAT GGC TTT GAC TTG AGA GGC TAC TTC TAT TGG TCT TTT ATG GAT AAC 380 381 Lys Ala Leu Asn Asp Gly Phe Asp Leu Arg Gly Tyr Phe Tyr Trp Ser Phe Het Asp Asn 1200 THE GAG TEG GET GAG CET TIT AGA CEA CEC TIT GEG CTG GTC GAG GTC GAC TAC ACG ACC 400 Phe Glu Trp Ala Glu Gly Phe Arg Pro Arg Phe Gly Leu Val Glu Val Asp Tyr Thr Thr TTC AAG ACG ACA CCG AGA AAG ACT GCT TAC ATA TAT GGA GAA ATT GCA AGG GAA AAG AAA Phe Lys Arg Arg Pro Arg Lys Ser Ala Tyr Ile Tyr Gly Glu Ile Ala Arg Glu Lys Lys 1120 ATA AAA GAC GAA CTG CTG GCA AAG TAT GGG CTT CCG GAG CTA TGA 440 1321 He Lys Asp Glu Leu Leu Ala Lys Tyr Gly Leu Pro Glu Leu End

Figure 6

THERMOCOCCUS CHITONOPHAGUS GLYCOSIDASE - 22G COMPLETE SEQUENCE - 9/95

COMPLETE SEQUENCE - 9/95
1 TTG CTT CCA COM
1 Het Leu Pro Glu Asn Phe Leu Trp Gly Val Ser Gln Ser Gly Phe Glu Phe Glu Het Gly 20
The Leu Trp Cly Val Ser Cly Bbc Cly Bb
61 CAC ACA CTC ACC
21 ASD AND LOW AND CAC ATT GAT CCA AAC ACA GAT TOO TOO
61 GAC AGA CTG AGG AGG CAC ATT GAT CCA AAC ACA GAT TGG TGG TAC TGG GTA AGA GAT GAA 120 121 ASP Arg Leu Arg Arg His Ile ASP Pro Asn Thr ASP Trp Trp Tyr Trp Val Arg Asp Glu 40
171 TAT AND
121 TAT AAT ATC AM AM GGA CTA GTA ACT CCC CAD
121 TAT AAT ATC AAA AAA GGA CTA GTA AGT GGG GAT CTT CCC GAA GAC GGT ATA AAT TCA TAT 180
41 Tyr Asn Ile Lys Lys Gly Leu Val Ser Gly Asp Leu Pro Glu Asp Gly Ile Asn Ser Tyr 60
181 GAA TTA TAT GAG AGA GAC CAA GAA GAC GAA GAA GAA GAA
181 GAA TTA TAT GAG AGA GAC CAA GAA ATT GCA AAG GAT TTA GGG CTC AAC ACA TAT AGG ATC 240 241 GGA ATT GAA TGG AGG AGA GTA TTA GGG ATG GBy Leu Asn Thr Tyr Arg Ile 80
and the fit He Ala Lys Asp Leu Gly Leu Aca TAT AGG ATC . 240
491 GGA ATT CAN man
81 Gly Ile Glu Trp Ser Arg Val Phe Pro Trp Pro Thr Thr Phe Val Asp Val Glu Tyr Glu 100
Ser Arg Val Phe Pro Trp Pro Thr Thr Phe Val Arg Val GAC GTG GAG TAT GAA 300
301 ATT CAT CAG CO.
301 ATT GAT GAG TCT TAC GGG TTG GTA AAG GAT GTG AAG ATT TCT AAA GAC GCA TTA GAA AAA 360 101 Ile Asp Glu Ser Tyr Gly Leu Val Lys Asp Val Lys Ile Ser Lys Asp Ala Leu Glu Lys 120 361 CTT GAT GAA ATC GCT AAG GLA AGA GAC GCA TTA GAA AAA 360
The ASP Glu Ser Tyr Gly Leu Val Lye Asp Vol. AAG ATT TCT AAA GAC GCA TTA CAA AAA
The val Lys Asp Val Lys Ile Ser Lys Asp Ala Lau Chan Ash 360
161 CIT GAT GAA ATC GCT AAC CAA ACC CA
161 CTT GAT GAA ATC GCT AAC CAA AGG GAA ATA ATA TAT TAT AGG AAC CTA ATA AAT TCC CTA 420
121 Leu Asp Glu Ile Ala Asp Glu Arg Glu Ile Ile Tyr Tyr Arg Asp Leu Ile Asp Ser Leu 140
421 AGA AAG AGG CGT TTT AAG GTA ATA CTA AAC CTA AAT CAT TTT ACC CTC CCA ATA TGG CTT 480
141 Arg Lys Arg Gly Phe Lys Val Ile Leu Asn Leu Asn His Phe Thr Leu Pro Ile Top Leu 160
Fire Lys Val Ile Leu Asn Leu Asn His Phe The The The The The The The The The T
481 CAT CAT COT and are
481 CAT GAT CCT ATC GAA TCT AGA GAA AAA GCC CTG ACC AAT AAG AGA AAC GGA TGG GTA AGC 540 161 His Asp Pro Ile Glu Ser Arg Glu Lys Ala Leu Thr Asn Lys Arg Asn Gly Trp Val Ser 180
FIG THE Glu Ser Arg Glu Lys Ala Leu Thr Ash AND AND AND THE GGA TGG GTA AGC 540
541 GAA AGG ACT COT AND
181 CON AGG ACT CIT ATA GAG TIT GCA AAA TIT COO COO
541 GAN AGG AGT GTT ATA GAG TIT GCN ANN TIT GCC GCG TAT TTA GCN TAT ANN TIT GGN GAC 600
181 Glu Arg Ser Val Ile Glu Phe Ala Lys Phe Ala Ala Tyr Leu Ala Tyr Lys Phe Gly Asp 200
601 ATA GTA GAC ATG TGG AGC ACA TTT AAT GAA CCT ATG GTG GTC GCC GAG TTG GGG TAT TTA 660
201 Ile Val Asp Het Trp Ser Thr Phe Ash Glu Pro Het Val Val Ala Glu Leu Gly Tyr Leu 220
Set in Phe Ash Glu Pro Met Val Ala Glu Leu Clu TAT TTA 660
661 GCC CCA TAC TCA GGA TTC CCC CCC CCC
661 GCC CCA TAC TCA GGA TTC CCC CCG GGA GTC ATG AAT CCA GAA GCA AAG TTA GTT ATG 221 Ala Pro Tyr Ser Gly Phe Pro Pro Gly Val Het Asn Pro Glu Ala Ala Lys Leu Val Het 240 721 CTA CAT ATG ATA AAG CCC CAR CCC CCG GGA GTC ANG TTA GTT ATG 720
The Pro Pro Cly Val Met Asn Pro Clu Ala Ala Lvo Tra GTT ATG 720
721 CTA CAT ATG ATA AAC GCC CAT GCT TTA GCA TAT AGG ATG ATA AAG AAA TTT GAC AGA AAA 780
241 Leu His Mer Tla And GCC CAT GCT TTA GCA TAT AGG ATG ATA AAG ANA
THE ASH ALS ALS LEU ALS THE MEE THE AND THE GAC AGA AAA 780
781 AAA GCT CAT COL 260
781 AAA GCT GAT CCA GAA TCA AAA GAA CCA GCT GAA ATA GGA ATT ATA TAC AAT AAC ATC GGC 840 261 Lys Ala Asp Pro Glu Ser Lys Glu Pro Ala Glu Ile Gly Ile Ile Tyr Asn Asn Ile Gly 280
ASP PER Glu Ser Lys Glu Pro Ala Glu Ile Gly Ile ANT TAC ANT AAC ATC GGC 840
841 GTC ACA THE GOA TH
841 GTC ACA TAT CCG TTT AAT CCG AAA GAC TCA AAG GAT CTA CAA GCA TCC GAT AAT GCC AAT 900 281 Val Thr Tyr Pro Phe Asn Pro Lys Asp Ser Lys Asp Leu Gln Ala Ser Lys Asp Leu Gln Ala Ser Lys Asp Ser Lys Asp Leu Gln Ala Ser Lys Asp Ser Lys Asp Leu Gln Ala Ser Lys Asp Lys Asp Leu Gln Ala Ser Lys Asp Lys Asp Leu Gln Ala Ser Lys Asp Ly
281 Val Thr Tyr Pro Phe Asn Pro Lys Asp Ser Lys Asp Leu Gln Ala Ser Asp Asn Ala Asn 300 901 TTC TTC CAC ACT CCC CTC TTC CAC ACT CCC AAT 300
901 TTC TTC CAC ACT AND ASS ASS ASS ASS ASS ASS ASS ASS ASS AS
THE THE CAC AGT GGG CTA THE THA AGE COTT AND DECEMBER OF THE THE THE AGE COTT AND DECEMBER OF THE THE THE AGE COTT AND DECEMBER OF THE THE THE THE THE THE AGE COTT AND DECEMBER OF THE
901 TTC TTC CAC AGT GGG CTA TTC TTA ACG GCT ATC CAC AGG GGA AAA TTA AAT ATC GAA TTT 960 961 GAC GGA GAG ACA TTT CTT GAG GGA GAG GGA GAG GAG ACA TTT GTG GAG GAG ACA TTT GAG GAG ACA TTT GTG GAG GAG ACA TTT GA
Det rie Leu Thr Ala Ile His Arg Gly Lys Leu Asn Ile Gly Pho 200
961 GAC GGA GAG ACA TIT GIT TAC CTT CCA TAT TTA AAG GGC AAT GAT TGG CTG GGA GTG AAT 1020
321 ASP Gly Glu Thr Phe Val Tri CCA TAT TTA AAG GGC AAT GAT TGG CTG GGA TTG
321 ASP Gly Glu Thr Phe Val Tyr Leu Pro Tyr Leu Lys Gly Asn Asp Trp Leu Gly Val Asn 340
1021 TAT TAT ACA ACA CAR CTA
341 TYP TYP THE ARE CAN AND COT AND THE COA ACT TYP COA ACT TYP COA ACT TYPE COA AC
ATY GIU VAI VAI LYS TYP GIN ASP PRO MET Phe Pro Ser ATE CCT CTC ATA 1080
1081 AGC TTC ANG CCC CTT 100
TALL COME OF COLD THE
SET Phe live Classics - TON TON OUR TUT ACE CON TON
Ser Phe Lys Gly Val Pro Asp Tyr Gly Tyr Gly Cys Ara CCA GGA ACG ACG TCA AAG GAC 1140
1081 AGC TTC AAG GGC GTT CCA GAT TAT GGA TAC GGA TGT AGA CCA GGA ACG ACG TCA AAG GAC 1140 361 Ser Phe Lys Gly Val Pro Asp Tyr Gly Tyr Gly Cys Arg Pro Gly Thr Thr Ser Lys Asp 380
1141 GGT AAT CCT GTT ACT COO 380
1141 GGT AAT CCT GTT ACT COO 380
1141 GGT AAT CCT GTT AGT GAC ATT GGA TGG GAG GTA TAT CCC AAA GGC ATG TAC GAC TCT ATA 1200 381 Gly Asn Pro Val Ser Asp Ile Gly Trp Glu Val Tyr Pro Lys Gly Mer Tyr Again 1200
1141 GGT AAT CCT GTT AGT GAC ATT GGA TGG GAG GTA TAT CCC AAA GGC ATG TAC GAC TCT ATA 1200 1201 GTA GCT GCC AAT GAA TAN TAN GCC ATG TAC GCC AAT GAA SGC ATG TAC GAC TCT ATA 400
1141 GGT AAT CCT GTT AGT GAC ATT GGA TGG GAG GTA TAT CCC AAA GGC ATG TAC GAC TCT ATA 1200 1201 GTA GCT GCC AAT GAA TAN TAN GCC ATG TAC GCC AAT GAA SGC ATG TAC GAC TCT ATA 400
1141 GGT AAT CCT GTT AGT GAC ATT GGA TGG GAG GTA TAT CCC AAA GGC ATG TAC GAC TCT ATA 1200 1201 GTA GCT GCC AAT GAA TAT GGA GTT CCT GTA TAC GTA ACA GAA AAC GGA ATA GGA GAT TCA 1201 Val Ala Ala Asn Glu Tyr Gly Val Pro Val Tyr Val Thr Glu Asn Glu Tac Gac GAT TCA 1260
1141 GGT AAT CCT GTT AGT GAC ATT GGA TGG GAG GTA TAT CCC AAA GGC ATG TAC GAC TCT ATA 1200 1201 GTA GCT GCC AAT GAA TAT GGA GTT CCT GTA TAC GTA ACA GAA AAC GGA ATA GCA GAT TCA 1260 1201 Val Ala Ala Asn Glu Tyr Gly Val Pro Val Tyr Val Thr Glu Asn Gly Ile Ala Asp Ser 1260 1261 AAA GAT GTA TTA ACC GCC TATA CCC TATA CCC AAA GGA AAC GGA ATA GCA GAT TCA 1260 1261 AAA GAT GTA TTA ACC GCC TATA CCC TATA CCC AAA GAA AAC GGA ATA GCA GAT TCA 1260
1141 GGT AAT CCT GTT AGT GAC ATT GGA TGG GAG GTA TAT CCC AAA GGC ATG TAC GAC TCT ATA 1200 1201 GTA GCT GCC AAT GAA TAT GGA GTT CCT GTA TAC GTA ACA GAA AAC GGA ATA GCA GAT TCA 1260 1201 Val Ala Ala Asn Glu Tyr Gly Val Pro Val Tyr Val Thr Glu Asn Gly Ile Ala Asp Ser 1260 1261 AAA GAT GTA TTA ACC GCC TATA CCC TATA CCC AAA GGA AAC GGA ATA GCA GAT TCA 1260 1261 AAA GAT GTA TTA ACC GCC TATA CCC TATA CCC AAA GAA AAC GGA ATA GCA GAT TCA 1260
1141 GGT AAT CCT GTT AGT GAC ATT GGA TGG GAG GTA TAT CCC AAA GGC ATG TAC GAC TCT ATA 1200 1201 GTA GCT GCC AAT GAA TAT GGA GTT CCT GTA TAC GTA ACA GAA AAC GGA ATA GCA GAT TCA 1260 1201 Val Ala Ala Asn Glu Tyr Gly Val Pro Val Tyr Val Thr Glu Asn Gly Ile Ala Asp Ser 1260 1261 AAA GAT GTA TTA ACC GCC TATA CCC TATA CCC AAA GGA AAC GGA ATA GCA GAT TCA 1260 1261 AAA GAT GTA TTA ACC GCC TATA CCC TATA CCC AAA GAA AAC GGA ATA GCA GAT TCA 1260
1141 GGT AAT CCT GTT AGT GAC ATT GGA TGG GAG GTA TAT CCC AAA GGC ATG TAC GAC TCT ATA 1200 1201 GTA GCT GCC AAT GAA TAT GGA GTT CCT GTA TAC GTA ACA GAA AAC GGA ATA GGA GAT TCA 1201 Val Ala Ala Asn Glu Tyr Gly Val Pro Val Tyr Val Thr Glu Asn Glu Tac Gac GAT TCA 1260

Figure 7a

1321	CLU A	T C	T TA	T GAC	Val	AGA Arg	GIY	TAC	TTA	CAC His	TCC	GCA	TTA	ACC	GAT	MT	TAC	GAA	TCC	Ling
																				460
1441	AN CC	C AG	G AAJ		. ~			·		.,.		•••	~==	reu	He	Thr	Lys	Clu	Ytd	1440 480
	Lys Pr									-			GTT Val	ATT Ile	AAT Asn	TAA TRA	GGG Gly	CTA Leu	ACA .	1500.
	AGC AA Ser As	CAT	CACC										36				·		,	. 500

Figure 7b(Continued)

PYROCOCCUS FURIOSUS GLYCOSIDASE - 7G1 - COMPLETE GENE SEQUENCE - 10/95

ATC TTC CONT.	
Het Phe Pro Glu Ive Pho CTI TGG GGT GTG GCA CAA TCG GGT TTG	
1 AIG TIC CCT GAA AAG TIC CIT IGG GGT GTG GCA CAA ICG GGT TIT CAG TIT GAA ATG GGG 61 GAT All COT THE PRO GLU LYS Phe Leu Irp Gly Val Ala Gln Ser Gly Phe Gln Phe Glu Ha	£Ο
	50
Asp Lys Leu Arg	20
61 GAT AAA CTC AGG AGG AAT ATT GAC ACT AAC ACT GAT TGG TGG CAC TGG GTA AGG GAT AAG 21 Asp Lys Leu Arg Arg Asn Ile Asp Thr Asn Thr Asp Trp Trp His Trp Val Arg Asp Lys 121 ACA AAT ATA GAG AAA GGC CTC GTT AGT GGC CAC TGG GTA AGG GAT AAG 121 ACA AAT ATA GAG AAA GGC CTC GTT AGT GGC CAC TGG GTA AGG GAT AAG 121 ACA AAT ATA GAG AAA GGC CTC GTT AGT GGC CAC TAGT TAGT	120
121 ACA AAT ATA GAG BER CCC and	40
41 The Ash Ile Glu Ive City GTT AGT GGA GAT CTT CCC CAG	40
121 ACA AAT ATA GAG AAA GGC CTC GTT AGT GGA GAT CTT CCC GAG GAG GGG ATT AAC AAT TAC 181 GAG CTT TAT GAG AAG GAC CAT GAG ATT GCA GAT CTC GCC GAG GAG GGG ATT AAC AAT TAC 181 GAG CTT TAT GAG AAG GAC CAT GAG ATT GCA GAG ATT GCA GAG GAG GAG GAG GAG GAG GAG ATT AAC AAT TAC 61 GIN IAN TATA GAG AAG GAC CAT GAG ATT GCA GAG ATT GCA ACC ATT GCA GAG ATT GCA ACC ATT GCA ACC ATT GCA ACC ACC ACC GAG ATT GCA ACC ACC ACC GAG ATT GCA ACC ACC ACC ACC ACC ACC ACC ACC ACC	100
181 GAG CTT TAT GAG AAG GAC CAR CAR CAR CAR CAR CAR CAR CAR CAR C	60
of Giu Lou Tyr Giu Lys Asp Car GCA ACA AAG CTG GGT CTT ASP	00
181 GAG CTT TAT GAG AAG GAC CAT GAG ATT GCA AGA AAG CTG GGT CTT AAT GCT TAC AGA ATA 61 Glu Leu Tyr Glu Lys Asp His Glu Ile Ala Arg Lys Leu Gly Leu Asn Ala Tyr 241 GGC ATA GAG TGG AGC AGA ATA TTC GCA TGG TGT TAC AGA ATA 81 Gly Ila Clu TGG AGC AGA ATA TTC GCA TGG TGT AGG TGT AGA ATA	240
241 GGC ATA GAG IGG AGG AGA ATA THE DOLLAR TO THE THE AGA ATA THE TYPE AND THE	80
241 GGC ATA GAG TGG AGC AGA ATA TTC CCA TGG CCA ACG ACA TTT ATT GAT GTT GAT TAT AGC 301 TAT AAT GAA TCA TAT AAC CTT ATA GAA GAT GTT TAT AFP Val Aap Val Aap Tyr Ser	
701 FIRE PEO TEP PEO The The Phe Tie ACT GAT TAT AGC	300
301 TAT AAT GAA TCA TAT AAC CTT ATA COL	100
301 TAT ANT GAN TON TAT AND CIT ATA GAN GAT GTN ANG ATC ACC ANG GAC ACT TTG GAG GAG 101 Tyr Ash Glu Ser Tyr Ash Leu Ile Glu Asp Val Lys Ile Thr Lys Asp Thr Leu Glu Glu 121 Leu Ash GND ATC GCC AND ANG AGG GAG GTC CCC and The Lys Asp Thr Leu Glu Glu	•
361 mm and Act TIG GAG GAG	360
121 1.4 GAT GAG ATC GCC AAC ANG ACC CAC	120
THE ASP GIU ILE ALE ASP LYS ASP CITY THE THE STORY AND ASP CITY	
361 TTA GAT GAG ATC GCC AAC AAG AGG GAG GTC GCC TAC TAT AGG TCA GTC ATA AAC AGC CTG 121 Leu Asp Glu lie Ala Asn Lys Arg Glu Val Ala Tyr Tyr Arg Ser Val lie Asn Ser Leu 421 AGG AGC AAG GGG TTT AAG GTT ATA GTT AND GTT	120
421 AGG AGG AAG GGG ITT AAG GTT ATA GTT AAT CTA AAT CAC TTC ACG CTT CCA TAT TGG ITG	140
141 Acr Ser Lys Gly Phe Lys Val lie Val Acr Leu Arn Mis Phe Thr Leu Pro Tyr Trp Leu 161 Acr CAT CCC ATT GAG GCT AGG GAG AGG GCG CCC	
481 Car cam one and the real real real real real real real rea	480
161 Min hal CCC ATT GAG GCT AGG GAG AGG GGG	160
48: CAT GAT CCC ATT GAG GCT AGG GAG AGG GCG TTA ACT AAT AAG AGG AAC GGC TGG GTT AAC 161 Kis Asp Pro 11e Glu Ala Arg Glu Arg Ala Leu Thr Ash Lys Arg Ash Gly Trp Val Ash 191 Pro Arg Aca GTT ATA GAG TTT GGA AAG TAT GGG AGG TTT GGA AAG TAT GGG AGG A	
541 CCA AGA ACA GTT ATA GAG TTT GCA AAG TAT GCC GCT TAC ATA GCC TAT AAG TIT GGA GAT 191 Pro Arg Thr Val Ile Glu Phe Ala Lys Tyr Ala Ala Tyr Ile Ala TYr Ile Ala TYR II GGA GAT	540
191 Pro Arg Thr Val 11e Glu Phe Ala Lys Tyr Ala Ala Tyr Ila Ala Tyr Lys Phe Gly Asp 601 ATA GTG GAT ATG TGG AGG ACG TTT AAT GAG GGT TAC ATA GCC TAT AAG TTT GGA GAT 201 Ila Val	180
THE VAL THE GIU Phe ALE LYS TWO ALL GOT TAC ATA GOO TAT ANG TIT CON COM	
	600
201 Ile Val ATT MGG AGC ACG TIT AAT GAG CCT ATC CON	200
601 ATA GTG GAT ATG TGG AGC ACG TTT AAT GAG CCT ATG GTG GTT GTT GAG CTT GGC TAC CTA 201 Ile Val Asp Met Trp Ser Thr Phe Ash Glu Pro Met Val Val Glu Leu Gly Tyr Leu 201 A's Pro TAC TCT GGC TTC CCT CCA GGG CTT GTT GTT GAG CTT GGC TAC CTA	
661 GCC CCC TAC TCT CCC	660 220
661 GCC CCC TAC TCT GGC TTC CCT CCA GGG GTT CTA AAT CCA GAG GCC GCA AAG CTG GCG ATA 721 CTT CAC ATG ATA AAT GCA CAT GCT TTA CCT TAC GAG GCC GCA AAG CTG GCG ATA 721 CTT CAC ATG ATA AAT GCA CAT GCT TTA CCT TAC GAG GCC GCA AAG CTG GCC ATA 721 Lau Nac Atta AAT GCA CAT GCT TTA CCT TAC GAG GCC GCA AAG CTG GCC ATA 721 Lau Nac Atta AAT GCA CAT GCT TTA CCT TAC GCT TAC	220
Day Pine Pro Pro Gly Val Leu Ash Bro Cha SCC GCA AAG CTG GCG ATA	720
721 CTT CAC ATG ATA ATT CC. CO. CO.	240
721 CTT CAC ATG ATA AAT GCA CAT GCT TTA GCT TAT AGG CAG ATA AAG AAG TTT GAC ACT GAG 241 Leu His Met Ile Asn Ala His Ala Leu Ala Tyr Arg Gln Ile Lys Lys Phe Asp Thr Glu 261 LW AND GAT AAG GAT TCT AAA GAG CCT CCA TATA AGG CAG ATA AAG AAG TTT GAC ACT GAG 261 LW AND GAT AAG GAT TCT AAA GAG CCT CCA TATA CCA TAT	• • •
Ala Leu Ala Tyr Arg Gln Ile Twa AAA TTT GAC ACT GAG	780
	260
781 AAA GCT GAT AAG GAT TCT AAA GAG CCT GCA GAA GTT GGT ATA ATT TAC AAC AAC ATT GGA 8 261 Lys Ala Asp Lys Asp Ser Lys Glu Pro Ala Glu Val Gly Ile Ile Tyr Ash Ash Ile Gly 2 261 GTT GCT TAT CCC AAG GAT CCG AAC GRT TGG 210	
841 GTT GCT TRE ARC ARC ATT CGA 8	40
	80
VAL ALA TYP PEO LYS ASD PEO ASD DAT TOO AND GAT CIT AND GCA GCA CAD AND DATE OF THE PEOPLE OF THE PE	
901 TTC TTC CAC men and Asia Call As	00
	90
301 Phe Phe His Ser Gly Leu Phe Phe Glu Ala Ile His Lys Gly Lys Leu Asn Ile Glu Phe 961 GAC GGT GAA ACG TTT ATA GAT GCC CGC TAT GTT ATA GAT GLU Phe 321 AND GLU GAA ACG TTT ATA GAT GCC CGC TAT GTT ATA GAT GCC CGC TATA GTT ATA	
961 GAC GGT CER AND ILLE GIVE THE MIS LYS GIV LYS LEU AND ILLE GIVE THE	60 -
321 ASD GIV GIV TO ATA GAT GCC CCC TAT CTD 330	20
961 GAC GGT GAA ACG TTT ATA GAT GCC CCC TAT CTA AAG GGC AAT GAC TGG ATA GGG GTT AAT 10 1021 TAC TAC ACA AGG GAA GTA GTT ACG TAT CAC TAT CAC TAT AAG TAT TAC TAC ACA AGG GAA GTA GTT ACG TAT CAC TAT CAC TAT TAC TAC ACA AGG GAA GTA GTT ACG TAT CAC TAT TAC TAC ACA AGG GAA GTA GTT ACG TAT CAC TAT TAC TAC TAC TAC TAC TAC TAC	220
	220
1021 TAC TAC ACA AGG GAA GTA GTT ACG TAT CAG GAA CCA ATG TTT CCT TCA ATC CCG CTG ATC 1081 ACC TTT AAG GGA GTT CAA GGA TAT GGC TAT CGC TAT CGC TAT GGA TAT GGC TAT GGG	10
ATG CIU Val Val Thr Tyr Gln Gln Dra CCA ATG TTT CCT TCA ATC CCG CTG ATC	080
361 The Phe Lys Gly Val Gln Gly Tyr Gly Tyr Ala Cys Arg Pro Gly The Leu Ser Lys Asp 381 Ann Arg CCC GTC AGC GAC ATA GGA TIG CAN GET GAS ARG PRO GLY THE Leu Ser Lys Asp 381 Ann Arg CCC GTC AGC GAC ATA GGA TIG CAN GET GAS AND ARG CCC GTC AGC GAC ATA GGA TIG CAN GET GAS AND ARG TIG GAS AND ARG TIG CAN GET GAS AND ARG TIG GAS AND ARG TI	
THE GIN GLY TYP GLY TYP ALA CVY AND DET GGA ACT CTG TCA AAG GAT 11	40
1141 GAC AGA CCC GTC AGC GAC ATA GGA TGG GAA CTC TAT CCA GAG GGG ATG TAC GAT TCA ATA 12	
ASP ILE GLY TEP Glu Leu TYP PER GLU CLG ATG TAC GAT TCA ATA 12	0C
401 Val Giu Ala His Lys Tyr Gly Val Pro Val Tyr Val Thr Giu Asn Gly Ile Ala Asp Ser 421	
TYP CITY VAL Pro Val Tyr Val Thr City and City GEG GAT TCA 12	60
421 ASH GLY ILE ALA Amp Ser 421)

Figure 8a

1261 421	Lys Asp					-						116	Lys	met	116	Glin	100	A: -	Db.	1320
	Glu Asp	Gly	TAI	€J ⊓ ŒYY	GIT Val	Lys	GGC. Gly	TAC Tyr	TTC Phe	CYC CYC	TGG Trp	GCA Ala	TTA Leu	ACT	GAC	AAC	TTC	GAG	TGG	1380 160
461	Ala Leu	C1y	Phe	AGA Arg	AIG Me I	yra CCC	TTT Phe	GGC Gly	CTC Leu	TAC Tyr	GAX Glu	STC Val	AAC Aan	CTA Leu	ATT	ACA	AAG	GAG	AGA	1440
481	Ile Pro	Arg	C) ri	Lys	NGC	GTG Val	TCG Ser	ATA Ile	TTC Phe	AGA Arg										1500
1201	AMA AAG	AIT	GAA	CAC	C 3 3							33					•			300

Figure 8b(Continued)

sankia gouldi endoglucanase (37071)

(370F1)
9 18 27 36
5' ATG AGA ATA CGT TTA COC 300 - 45 54
Met Arg Ile Arg Leu Ale The Cit GCG CTC TGC GCA GCG CTG AGC CCA GCC AGC
Met Arg Ile Arg Leu Ala Thr Leu Ala Leu Cys Ala Ala Leu Ser Pro Val Thr
D.3 ' ••
TTT CCA GAT AAT CTA ACT
Phe Ala ASD AND Value CTA CAA ATC GAC GCC GAC GCC GGT AND AND COMPANY OF THE COMPANY OF THE CASE OF TH
Phe Ala Asp Asn Val Thr Val Glm Ile Asp Ala Asp Gly Cly Lys Lys Leu Ile
AGC CGA GCC CTT TAC GGC ATG AAT AAC TCC AAC GCA CAA AGC CTT ACC GAT ACT Ser Arg Ala Leu Tyr Gly Met Asn Asn Ser Asn Ala Gly Ser ACT
Ser Arg Ala Louis GGC ATG ART AAC TCC AAC GCA CAA AGC CTM 162
Ser Arg Ala Leu Tyr Gly Met Asn Asn Ser Asn Ala Glu Ser Leu Thr Asp Thr
171 180 171 180
GAC TYPE CAP COM 180 189 198 200
CAC TGG CAG CGT TIT CGC GAT GCA GGT GTG CGC ATG CTG CGG GAA AAT GGC GGC ABP Trp Gln Arg Phe Arg Asp Ala Gly Val Arg Met Leu Arg GA AAT GGC GGC
The Gir Arg Phe Arg Asp Ala Cly Val Arg Mor Louis Cod GAA AAT GGC GGC
And Giv Sin Ash Giv Giv
443 774
700 AGC ACC 123 mm 261 261
ABN ABN SET THE LYS TYP ASH TER GIR LET HE ARE AGT CAT CCG GAT TGG
The set us bid was the
4/3 200
THE AAU ANT GTC TAC COO COO COO
TYT Asn Asn Val TyT Ala Gly Asn Asn Asn Too GAC AAC CGG GTA GCC CTG ATT
Tyr Asn Asn Val Tyr Ala Gly Asn Asn Asn Trp Asp Asn Arg Val Ala Leu Ile
and GVV VVC CALC CALC CALC CALC CALC CALC CAL
CAG GAA AAC CTG CCC GGC GCC GAC ACC ATG TGG GCA TTC CAG CTC ATC GGT AAG Gln Glu Asn Leu Pro Gly Ala Asp Thr Net Trp Ala Pho Clar CTC ATC GGT AAG
Gln Glu Asn Leu Pro Gly Ala Asp Thr Met Trp Ala Phe Gln Leu Ile Gly Lys
30/ 76c
OLU GCG ACT TYPE CON
Val Ala Ala Thr Ser Ala Tyr hen Bha had GAT TGG GAA TTC AAC CAG TGG CAA
Val Ala Ala Thr Ser Ala Tyr Asn Phe Asn Asp Try Glu Phe Asn Gln Ser Gln
741 150
TGG TGG ACC GCC GCC GCC GCC ACC ACC ACC ACC A
Trp Trp Thr Gly Val Ala Gln Asn Leu Ala Gly Gly Gly Glu Pro Asn Leu Asp
Ash Leu Ala Cly Gly Gly Pro Ash Leu Ash
973 664
GGC GGC GAA COT COT 522 531 531
Gly Gly Glu Ala Leu Val Glu Gly Asp Pro Ash Leu Tyr Leu Het Asp Trp
val Ciu Cly Asp Pro Asn Leu Tyr Leu Mer Asn Ton
TCG CCA GCC GAC ACM GCC GAC
Ser Pro Ala ABD Thr Wal City ATT CTC GAC CAC TGG TTT GGC GTA ABC CTC
Ser Pro Ala Asp Thr Val Gly He Leu Asp His Trp Phe Gly Val Asn Gly Leu
943 219
GCC GTG CGG CGT CCC 112 621 630 639
Gly Val Arg Arg Gly Live ANA TAC TOG AGT ATG GAT AAC GAG COG
Gly Val Arg Arg Gly Lys Ala Lys Tyr Trp Ser Net Asp Asn Glu Pro Gly Ile
TGG GTT GGC ACC GGG 675 684 593
TGG GTT GGC ACC CAC GAC GAT GTA GTG AAA GAA CAA ACG CCG GTA GAA GAT TTC TTP Val Gly Thr His Asp Asp Val Val Lys Gly Gly Thr Dec Val GAA GAT TTC
Trp Val Gly Thr His Asp Asp Val Val Lys Glu Gln Thr Pro Val Glu Asp Phe
wat Glu Asp Phe
The state of the s

Figure 9a

Bankia gouldi endoglucanese (37071) (continued)

720 CTG CAC ACC TAT TTC GAA ACC GCC AAA AAA GCC CGC GCC AAA TTT CCC GGT ATT 729 Leu His Thr Tyr Phe Glu Thr Ala Lys Lys Ala Arg Ala Lys Phe Pro Gly Ile 783 -ANA ATC ACC GGT CCG GTG CCC GCT AAT GAG TGG CAG TGG TAT GCC TGG GGC GGT Lys Ile Thr Gly Pro Val Pro Ala Asn Glu Trp Gln Trp Tyr Ala Trp Gly Gly 828 TTC TCG GTA CCC CAG GAA CAA GGG TTT ATG AGC TGG ATG GAG TAT TTC ATC AAG Phe Ser Val Pro Gln Glu Gln Gly Phe Met Ser Trp Met Glu Tyr Phe Ile Lyr 882 CGG GTG TCT GAA GAG CAA CGC GCA AGT GGT GTT CGC CTC CTC GAT GTA CTC GAT 891 Arg Val Ser Glu Glu Gln Arg Ala Ser Gly Val Arg Leu Leu Asp Val Leu Asp 927 936 CTG CAC TAC TAC CCC GGC GCT TAC AAT GCG GAA GAT ATC GTG CAA TTA CAT CGC Leu His Tyr Tyr Pro Gly Ala Tyr Asn Ala Glu Asp Ile Val Gln Leu His Arg 990 ACG TTC TTC GAC CGC GAC TTT GTT TCA CTG GAT GCC AAC GGG GTG AAA ATG GTA 999 Thr Phe Phe Asp Arg Asp Phe Val Ser Leu Asp Ala Asn Gly Val Lym Met Val 1035 GAA GGT GGC TGG GAT GAC AGC ATC AAC AAG GAA TAT ATT TTC GGG CGA GTG AAC 1053 Glu Gly Gly Trp Asp Asp Ser Ile Asn Lys Glu Tyr Ile Phe Gly Arg Val Asn 1098 1107 GAT TOG CTC GAG GAA TAT ATG GGG CCA GAC CAT GGT GTA ACC CTG GGC TTA ACC Asp Trp Leu Glu Glu Tyr Met Gly Pro Asp His Gly Val Thr Leu Gly Leu Thr 1152 GAA ATG TGC GTG CGC AAT GTG AAT CCG ATG ACT ACC GCC ATC TGG TAT GCC TCC 1161 Glu Met Cys Val Arg Asn Val Arn Pro Met Thr Thr Ala Ile Trp Tyr Ala Ser 1206 1215 ATG CTC GGC ACC TTC GCG GAT AAC GGC GTC GAA ATA TTC ACC CCA TGG TGC TGG Met Leu Gly Thr Phe Ala Asp Asn Gly Val Glu Ile Phe Thr Pro Trp Cys Trp 1251 1260 AAC ACC GGA ATG TGG GAA ACA CTC CAC CTC TTC AGC CGC TAC AAA CCT TAT Asn Thr Gly Met Trp Glu Thr Leu His Leu Phe Ser Arg Tyr Asn Lys Pro Tyr 1314 1323 CGG GTC GCC TCC AGC TCC AGT CTT GAA GAG TTT GTC AGC GCC TAC AGC TCC ATT Arg Val Ala Ser Ser Ser Ser Leu Glu Glu Phe Val Ser Ala Tyr Ser Ser Ile ANC GAN GCA GAN GAC GCC ATG ACG GTA CTT CTG GTG AAT CGT TCC ACT AGC GAC 1377 Asn Glu Ala Glu Asp Ala Met Thr Val Leu Leu Val Asn Arg Ser Thr Ser Glu

Figure 9b(Continued)

Bankia gouldi endoglucanase (370P1) (continued)

ACC CAC ACC GCC ACT GTC GCT ATC GAC GAT TTC CCA CTG GAT GGC CCC TAC CGC Thr His Thr Ala Thr Val Ala Ile Asp Asp Phe Pro Lou Asp Gly Pro Tyr Arg

1467 1476 1485 1494 1503 1512
ACC CTG CGC TTA CAC AAC CTG CCG GGG GAG GAA ACC TTC GTA TCT CAC CGA GAC
Thr Leu Arg Leu His Asn Leu Pro Gly Glu Glu Thr Phe Val Ser His Arg Asp

1521 1530 1539 1548 1557 1566
AAC GCC CTG GAA AAA GGT ACA GTG CGC GCC AGC GAC AAT ACG GTA ACA CTG CAG
AEn Ala Leu Glu Lys Gly Thr Val Arg Ala Ser Aep Aen Thr Val Thr Leu Glu

1575 1584 1593 1602 1611
TTG CCC CCT CTG TCC GTT ACT GCA ATA TTG CTC AAG GCC CGG CCC TAA 3'
Leu Pro Pro Leu Ser Val Thr Ala Ile Leu Leu Lys Ala Arg Pro ***

Pigure 94 (Continued)

Thermotoga maritima Alpha-oalactosidane Complete Gene Sequence (1 c + 3)

5' GTG ATC TGT GTG GAA ATA TITC GGA AAG ACC TTC AGA GAA GAA GAA TTC GTT CTC
Val Ile Cys Val Glu Ile Phe Gly Lys Thr Phe Arg Glu Gly Arg Phe Val Leu
63 72
AN GAG AN AND 37TH ACA (TITLE GAG 81 90 99 100
ANA GAG ANA AND THE ACA CIT GAG THE GCG GTG GAG ANG ATA CAC CIT GGC TGG
Lys Glu Lys Asn Phe Thr Val Glu Phe Ala Val Glu Lys Ile His Leu Gly Trp
11/ 996
ANG ATC TCC GGC AGG GTG ANG GGA AGT CCG GGA AGG CTT GAG GTT CTT CGA AGG
Lys Ile Ser Civ. Aug.
Lys Ile Ser Gly Arg Val Lys Gly Ser Pro Gly Arg Leu Glu Val Leu Arg Thr
1/1 100 -
ANA GCA CCG GAA ANG GTA CTT GTG ANC ANC TOG CAG TOC TOG GGA CCG TGC AGG
Lys Ala Pro Glu Lys Val Leu Val Asn Asn Trp Cln Ser Trp Gly Pro Cys Arg
GTG GTC GAT GCC TTT TCT TTC A33 252 261 270
THE ANA COA COT GAA ATA GAT COG AAC TGG AGA TAC
Val Val Asp Ala Phe Ser Phe Lys Pro Pro Glu Ile Asp Pro Asm Trp Ary Tyr
279 200 200
ACC GCT TCG GTG GTG CCC GAT GTA CTT GAA AGG AAC CTC CAG AGC GAC TAT TTC
Thr Ala Ser Val Val Pro Asp Val Lou Clu Annual
Thr Ala Ser Val Val Pro Asp Val Leu Glu Ary Asm Leu Glm Ser Asp Tyr Phe
GTG GCT GAA GAA GGA AAA GTG 351 360 369 378
GTG GCT GAA GAA GGG TAC GGT TIT CTG AAT ATC GCA CAT CCT
Val Ala Glu Glu Gly Lys Val Tyr Gly Phe Leu Ser Ser Lys Ile Ala His Pro
387 206
THE THE GET GIG GAA GAT GGG GAA CIT GIG GCA TAC CITE GAA TAT THE GAT GIC
Phe Phe Ala Val Glu Aco Clu Cou Val
Phe Phe Ala Val Glu Asp Gly Glu Leu Val Ala Tyr Leu Glu Tyr Phe Asp Val
GAG TTC GAC GAC TTC GTT GTT GTT GTT GAC GAC TTC GAC GAC TTC GTT GTT GTT GTT GTT GTT GTT GTT GT
GAG TTC GAC GAC TTT GTT CCT CTT GAA CCT CTC GTT GTA CTC GAG GAT CCC AAC
Glu Phe Amp Asp Phe Val Pro Leu Glu Pro Leu Val Val Leu Glu Asp Pro Asm
495 604 605
ACA CCC CIT CTT CTG GAG AAA TAC GCG GAA CTC GTC GGA ATG GAA AAC AAC GCG
The Pro Leu Leu Clu Lyr The Ala Cl
The Pro Leu Leu Glu Lys Tyr Ala Glu Leu Val Gly Met Glu Asn Asn Ala
549 558 567 576 585 594
AGA GTT CYA ANA CAC ACA CCC ACT CGA TCG TCC AGC TCG TAC CAT TAC TTC CTT
Arg Val Pro Lys His Thr Pro The Gly Trp Cys Ser Trp Tyr His Tyr Phe Leu

Figure 10a

Thermotoga maritima Alpha-galactosidade Complete Gene Sequence (2 01 1)

(
GAT CTC ACC TOG GAA GAG ACT CTC AAG AAC CTC AAG CTC OCG AAG AAT TTC CCC ASP Leu Thr Trp Glu Glu Thr Leu Lys Asn Leu Lys Leu Ala Lys Ann Phe Pro
657 666 675 684 693 702 TTC GAG GTC TTC CAG ATA GAC GAC GCC TAC GAA AAG CAC ATA GGT GAC TGG CTC Phe Glu Val Phe Gln Ile Asp Asp Ala Tyr Glu Lys Asp Ile Gly Asp Trp Leu
OTG ACA AGA GGA GAC TIT CCA TCG GTG GAA GAG ATG GCA AAA GTT ATA CCG GAA Val Thr Arg Gly Asp Phe Pro Ser Val Glu Glu Met Ala Lys Val Ile Ala Glu
AMC COT TIC ATC CCG GGC ATA TGG ACC GCC CCG TIC AGT GTT TCT GAA ACC TCG ASD Gly Phe Ile Pro Gly Ile Trp Thr Ala Pro Phe Ser Val Ser Glu Thr Ser
GAT GTA TTC AAC GAA CAT CCG CAC TGG GTA GTG AAG GAA AAC GGA GAG CCG AAG ASP Val Phe Asm Glu His Pro Asp Trp Val Val Lys Glu Asm Gly Glu Pro Lys
ATG GCT TAC AGA AAC TGG AAC AAA AAG ATA TAC GCC CTC GAT CTT TGG AAA GAT Met Ala Tyr Arg Asn Trp Asn Lys Lys Ile Tyr Ala Leu Asp Leu Ser Lys Asp
CAG GTT CTG AAC TOG CTT TTC GAT CTC TTC TCA TCT CTG AGA AAG ATG GGC TAC Glu Val Leu Asn Trp Leu Phe Asp Leu Phe Ser Ser Leu Arg Lys Met Gly Tyr
AGG TAC TITC AAG ATC GAC TIT CTC TTC GCC GGT GCC GTT CCA GGA GAA AGA AAA Arg Tyr Phe Lys Ile Asp Phe Leu Phe Ala Gly Ala Val Pro Gly Glu Arg Lys
AG AAC ATA ACA CCA ATT CAG CCG TTC AGA AAA CGG ATT GAG ACG ATC AGA AAA LYS ASI lle Thr Pro Ile Glin Ala Phe Arg Lys Gly Ile Glu Thr Ile Arg Lys
La Val Gly Glu Asp Ser Phe Ile Leu Gly Cys Gly Ser Pro Leu Leu Pro Ala
1143 1152 1161 1170 1179 1188 TG GCA TGC GTC GAC GCG ATG AGG ATA GGA CCT GAC ACT GCG CCG TTC TGG GGA al Gly Cys Val Asp Gly Met Arg Ile Gly Pro Asp Thr Ala Pro Phe Trp Gly

Figure 10 (Continued)

Thermotoga maritima Alpha-calactosidade Complete Gene Sequenca (3.51.5)

1197 GAA CAT ATA	1206	1215	1224	1233	1242
			ce our our	VCV 100 CCC	CTG AGA AAC. CO
Glu His Ile	Glu Asp Asn	Cly Ala P	ro Ala Ala	ATE TTP Ala	ren yra yzu yr
1251	1200				
Ile The har	THE TIC AIG	CAC GAC A	∞ mc ∞ (TTG AAC GAC	287 1290 CCC CAC TOT CTC
112	AAT PUG Met	His Asp A	g Phe Trp 1	en yeu yeb	Pro Asp Cys Leu
ATA CTG AGA	CAG CAG AAA	ACG GAT CT	13 XCX CXG X	32 13 AG GAA AAG 0	341 1350
Ile Leu Arg	Glu Glu Lys	Thr Asp Le	u Thr Gln L	ys Glu Lva G	ilu Leu Dr Ser
1134	1170				
	ACY CIC CIC		A SEACE VITY OF	WA ACC GAT G	AT CITY THE CITY
-34 ALL CYS (My Val Leu	yzb yzu Hei	Ile Ile G	lu Ser Asp A	sp Leu Ser Leu
7077	7 / 7 7				
Val Arg Asp H	is Cly to a			<u> </u>	C CALC CAM CAN
1457		As Agt red	Lys Glu Th	r Leu Glu Le	Leu Gly Gly
YCY CCY CCC C	1476 TT CAA AAC A	1485 TC ATG TCG	GAG GAT CT	4 150	G ATC GTC TCG
Ary Pro Ary V	sl Gln Asn I	le Met Ser	Glu Asp I		a vic eic ice
1321	1570				-
TCT GGC ACT CT		- 0.0 7000	WIC GIG GIG	GAT CTG AM	TACC ACT COS
Ser Gly Thr Le	u Ser Gly A	sn Val Lys	Ile Val Val	. Asp Lee 26.	Car Lua Gin
· 13/3	1004				
TAC CAC CTG GA		y yag icc	TCC CTG AAA	YYY YCY GIG	GTC AAA AGA
TYT His Leu Gl	r TAR GIA CJ	y Lys Ser	Ser Leu Lys	Lys Arg Val	Val Lys Arg
AA CAC GCA AG	1638 AAC TTC TA	1647 TTC TAC (1656 EVA GAG GCT	1665 CAG AGA GAA	TCA 3
ilu Asp Gly Arg	Asn Phe Ty	Phe Tyr	lu Glu Gly	Glu Arg Glu	

Figure 10c(Continued)

Thermotoga maritima β -mannanase (Sept.)

			9			18									45			54
5,	ATG	GGG	ATT	GGT	GGC	CYC	GAC	TCC	TGG	AGC	CCG	TCA	GTA	TCG	CCG	GAA	TTC	CIT
	Met	GIA	He	GIA	GIĀ	YSD	Asp	Ser	Trp	5er	Pro	Ser	Val	Ser	λla	Glu	Phe	Leu
			63		•	72			81			90			99			10B
	TTA	TTG	ATC	GTT	GAG	CTC	TCT	TTC		CTC	TTT		AGT	CAC		TTC	GTG	777
	Leu	Leu	Ile	Val	Glu	Leu	Ser	Phe	Val	Leu	Phe	Yla	Ser	Asp	Glu	Phe	Val	Lys
	GTG	GAA	117	CCA	111	126	ححت	CTG	135	CCA		144	بكلمك	BCB	153			162
																AIT		AGC
	Val	Glu	Asn	Gly	Lys	Phe	Ala	Leu	λsn	Gly	Lys	Glu	Phe	Arg	Phe	Ile	Gly	Ser
																	_	
			171			180			189			198			207			
		AAC	TAC	170	V10	CAC	TAC	λAG	AGC	AAC	GUA	AIG	ATA	GAC	AGT	GTT	CIG	GAG
	λsn	Asn	TYI	Tyr	Het	His	TYI	Lys	Ser	Asn	Gly	Met	lle	λερ	Ser	Val	Leu	Glu
			_	_			_	_			-			•				
			225		•	234			243			252			261			270
	AGT	CCC	λGλ	GAC	ATG	GGT	ATA	λλG	CLC	circ	УСУ	ATC	TGG	CCI	TTC	CIC	CYC	CCC
	Ser	Ala	Ara	ASD	Met	GIV	Tle	Lys	Val.	Ten	Ara	710		Gly	Pho	Lou	>	C1:-
			,	,		U -3.		<i>ک</i> ړی			<i>,</i>		ν	Cly	- 111	Dea	ىرىم	GIA
			279			288			297			306			315			324
	GYC	AGT	TAC	TGC	λGλ	CYC	λAG	AAC	YCC	TAC	ATG	CAT	CCI	CYC	ccc	CCI	CII	TTC
	Gly		~~~	~	1)		\		~~~	 War		P			C)		
	514	SET	TYL	Cys	AL 9	رومم	pys	ASII	1111	ıyı	nec	ura	PLU	GIU	PTO	GIĀ	ATT	Phe
			333			342			351			360			369	-		378
	GGG	CIC	CCY	GΑλ	GGY	ATA	TCG	YYC	GCC	CAG	AGC	GGT	TIC	GAA	AGA	CIC	CAC	TAC
	GIY	VAI	PTO	GIU	GIA	17.0	Ser	Asn	ALB	GIN	Ser	GIA	PDe	Glu	Arg	Leu	yab	Tyr
			387			396			405			414			423			432
	ACA	CII	CCC	XXX	GCG	λλλ	CAA	CIC	GGT	λτλ	λλλ	CTT	GTC	λTT	GII	CIT	GTG	
	Thr	Val	YIS	Lys	YTA	Lys	Glu	Leu	Gly	Ile	Lys	Leu	Val	Ile	Val	Leu	Val	Asn
			441			450	•		459		,	468			477			486
	AAC	TGG			TTC		GGA	ATG		CAG	TAC			TGG			GGA	
	Asn	Trp	уер	Asp	Phe	GJĀ	Gly	Met	λsn	Gln	Tyr	Val	Arg	Trp	Phe	Gly	CJA	Thr
			495			504			513			522			531			540
	CAT	CAC			TTC		AGA	CAT			ATC			GAG			AAG	TAC
	Ris	His	Asp	Asp	Phe	Tyr	λrg	Asp	Glu	Lys	Ile	Lys	Glu	Glu	TVI	Lvs	Lvs	Tyr

Figure 11a

		•			•													
				oga	MAX	1112	na f	3-ma:	nnan		(334	-	(=	onti	nue	a) (رس ري	イン)
		54			55	8		56	7	•	57	6	•	58	E			
GT	c rc	CT	L CI	C CI	<mark>አ አ</mark> አ	C CA	T GT	C 77	TAC	C TA	CAC	G GG	. CT	םנ ייים יו	יים אינה ביים	c	594 G GAJ	
																C AG	G GAJ	
Va.	ı Se			u Va	l As:	n Hi	s Va	l As	n Th	r Ty	r Th	r Gly	y Val	l Pr	o Ty	r Ar	g Glu	
		60	3		61	2		62	1		63	0		63	٥			
GAG	i CC	C YC	C AT	CAT	G GC	CIC	G GA	G CT	T GC	A AAG	C GA	A CCC	CCC	TG	r GN	C >~	648 G	
61.		~	 :													- AC	S GAC	•
011	A PI	o m	I II	e Me	E Ala	a Tr	p Gl	u Le	n YJ	a Ası	a Gli	a Pro	Arg	Cys	5 G1:	ים ליים		
		65												_			ر دم.	,
AAZ	TC			C 30	660 	b 		67	5		6 B 4	1		693)		702	
						- GT	r GA	G TG	2 CI	G AAC	GAC	3 ATC	AGC	700	TAC	C ATA	702 እ አአር	
Lys	Se	r Gl	v Ası	ימים ב	r Lei		1 61.											
	•					. va.	L GI	ת יודן	P Va.	LLYS	.Glt	1 Met	Ser	Sex	Typ	: Ile	 : Lys	
	-	71:	1	•	720)		726			~~~							
AGT	, CIK	GA:	r cc	AAC	CAC	CIC	: GI	GC	. פוע	: 000	730 3 C A C	י רכא א	~~	747	·		756 AAC	
Ser	Let	ysI	Pro) Asi	His	Lev	va]	LAla	Va]	GIV	, year	Glu	GIV	Dh-			λsn	
										4				FILE	Pne	: Ser	\ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \	
T) C		765			774	l		783	1		792	:		801			810	
IAC	- GA.	ا لانک	TTC	: אאא	CCI	TAC	: cc:	CCX	GAJ	CCC	CAG	TGG	GCC	TAC	AAC	GGC	810 TGG	
- , -		. 623	7,116	. Lys	PTO	TYI	Gly	Gly	Glu	: Ala	Glu	Trp	λla	Tyr	Asn	Gly	Trp	
		819			828			837								_	-	
TCC	GGT			TGG	AAG	AAG	. (93/	~~	2005	846			855			864 ACG	
									100	AIA	GAG	ACG	GIG	GAC	TTC	CCC	XCG	
Ser	Gly	Val	λsp	Trp	Lys	Lys	Leu	Leu	Set	Tie	63	Thr						
			-		_	•				+10	GIU	1111	Agi	Asp	Phe	Gly	Thr	-
		873			882			891			900			909				
TTC	CAC	CIC	TAT	CCC	TCC	CAC	TCC	GGT	GTC	AGT	CCA	GAG	AAC	TET			318	
FIIE	nis	Leu	TYT	Pro	Ser	His	Trp	Gly	Val	Ser	Pro	Clu	Asn	Tyr	λla	Gln	T	
		927			936								-	_				
GGA	GCG			ልጥል	270	C3.C	~~~	945			954			963			972	
				717		GAC	CAC	ATA	AAG	ATC	GCA	λλλ	GAG	ATC	GGA	AAA	CCC	
		-	•			4	1113	116	LYS	118	VTS	ГЛЗ	Glu	Ile	Cly	Lys	Pro	
		981			990			999		1	800							
GTT	GTT	CTG	GAA	GAA	TAT	GGA	ATT	CCA	AAG	AGT	CCC	CCA		017		1	.026	
													211	AAC	AGA	ACG	CCC	
Val	Val	Leu	Glu	Glu	Tyr	Gly	Ile	Pro	Lys	Ser	λla	Pro	Val	A		M b	Ala	
									-						w. g	Inr	WTS .	
A T~		1035	~	1	044		3	053		1	062		1	071		1	080	
A.C	1 AC	AUA	CIC	TGG	YYC	GAT	CTG	GTC	TAC	GAT	CIC	CCT	GGA	GAT	GGA	GCG	ATG	
	- 7 -	~L 9	₽	TID	ASD.	ASD	Leu	Val	Tyr	ABD	Leu	Gly	Gly	λsp	Glv	Ala	Met	

Figure 11b(Continued)

Thermotoga maritima β-mannanase (mac) (continued) (6 CP2)
1089 1089
TTC TGG ATG CTC CCC 11078 1116
Phe Trp Met Leu Ala Gly Tla Clu Cl
1143 Asp Glu Arg Gly Tyr
TAT CCG GAC TAC CAS STEE 1161 1170 1170
TAT CCG GAC TAC GAC GGT TTC AGA ATA GTG AAC GAC GAC AGT CCA GAA GCG GAA TYT Pro Asp Tyl Asp Gly Plan
TYT Pro Asp Tyr Asp Cly Ph
Tyr Pro Asp Tyr Asp Gly Phe Arg Ile Val Asn Asp Asp Ser Pro Glu Ala Glu
1197 1206 1215 1774
TIG ATA AGA GAA TAC CCC AAC CCC TAC CC
1242
Leu Ile Arg Glu Tyr Ala Lys Leu Phe Asn Thr Gly Glu Asn
125 ASP IIe Arg Glu Asp
1251 1260 1269 1278
ACC TGC TCT TTC ATC CTT CCA AAA GAC GGC ATG GAG ATC AAA AAG ACC GTG GAA Thr Cys Ser Phe Ile Law
The Cys Ser Phe Ile Leu Pro Lys Asp Clu Nos Con And And Acc GTG GAA
Thr Cys Ser Phe Ile Leu Pro Lys Asp Gly Het Glu Ile Lys Lys Thr Val Glu
1305 1314 1323 1332
Val Arg Ala Gly Val Phe Asp Tyr Ser Asn Thr Phe Gly
Val Arg Ala Gly Val Phe Asp Tyr Ser Asn Thr Phe Glu Lys Leu Ser Val Lys
1359 1360
GTC GAA GAT CTG GTT TTT GAA AAT GAG ATA GAG CAT CTC GGA TAC GGA ATT TAC Val Glu AED Leu Val Die en
THE TAX AAT GAG ATA GAG CAT CTC GGA TAC GGA ATT TAC
Val Glu Asp Leu Val Phe Glu Asn Glu Ile Glu Bir Lou Gl
and bed Gly Tyr Gly Tie Tom
1413 1422 1431 1440 1449 1469
ACC CGG ATC CCG GAT GGA GAA CAM GAS
Gly Phe Asp Leu Asp Thr Thr Arg Ile Pro Asp Cly Cly Pi
Gly Phe Asp Leu Asp Thr Thr Arg Ile Pro Asp Gly Glu His Glu Met Phe Leu
1467 1476 1485
GAA GGC CAC TIT CAG GGA AAA ACG GTG AAA CAG MEE 1503 1512
GAA GGC CAC TIT CAG GGA AAA ACG GTG AAA GAC TCT ATC AAA GCG AAA GTG GTG Glu Gly His Phe Gln Gly Lys Thr Val Lys A
Gly His Phe Gln Gly Lys Thr Val Lys Asp Ser Ile
Glu Gly His Phe Gln Gly Lys Thr Val Lys Asp Ser Ile Lys Ala Lys Val Val
AAC GAA GCA CGG TAC 550 1539 1548 1557
CAC GAG GAA GTT CAT TITE TO 1566
Asn Glu Ala Arg Tyr Val Leu Ala Glu Chu th
THE SET SET DESCRIPTION
1575 1584 1593 1602 1611 1520
THE WAR ALL TEG CAG CCA CAC MEDICAL TOPON
ALL LYS ASH TED TED ASS SON ON
'al Lys Asn Trp Trp Asn Ser Gly Thr Trp Gln Ala Glu Phe Gly Ser Pro Asp

Figure 11C(Continued)

		The	PRO	eog:		ar:	مدعا	44	β- -	L D.D.	LII a		Œ.	(E)		cont	inne	.a.	(60	Ξρ,
AT	T GJ	16	29		1	626													_	
71	 e G]	u T	 πο λ	 sn (31 v					AT 0		GC	·			166 G AA		אג ס בב	y CI	Ċ.
		705	13		•	600										u As				
ccc	GG 	A A2					· un	, W	M G	IG Y	GŊ.	GTA	, ec	O A AG	G AA	G TT	9 C Ga	A AG	172	8
Pro	G1	y Ly	'S 50	er A	Sp	TIP	Gl	a G1	u V	al A	rg	Val	A1	a Ar	g Ly	s Ph	 • Gl:	 u Ar	 G T.e	-
		1/1	,		•	710										177; C GIY				
Ser	Gl	r cy	3 G1	u I	le i	Leu	Glu	ı Ty	z ya	p I	le	Tyr	Ile	Pro	As:	n Val	Gli	 : Gl ₂	· Cro · Let	-
•		179	1		16	300										1827 GTG				
Lys	GJ7	/ Ar	J Le	נע מ	rg F)TO	Tyr	A1	va	l Le	- :u	Asn	Pro	G17	TI	Val	Lys	Ile	Gly	•
		1845	5		1 9	5.4				_										
Leu	Asp	Met	λsı	a As	n A	la	λsn	Val	Gli	se	 :	Ala	 Glu	Ile	Ile	Thr	Phe	Gly	Gly	
		1899	1		10	ΛP				_						1935 ACA				
Lys	Glu	TYI	λrc	λr	g P	be :	His	Val	λrg	111	 e 0	lu	Phe	λsp	Arg	Thr	λla	Gly	 Val	
		1953			194	62										1989 Gat				
Lys	Clu	Leu	His	Il	: e G]	ly i	/al	 Val	Gly	, vai	 > H	is i	 Leu	Arg	Tyr	Asp	Gly	 Pro	ATT	
	- 2	2007			201	_														
TTC	ATC	GAT 	AAT 	GIC	• AU		-1-T	TAT	AAA	УСУ	A	CA (GA	CCT	atg	2043 TGA	3 '			
Phe :	Ile	Авр	λsn	Val	l Ar	g L	æu '	ŢYI	Lys	λrg	T	pr c	ЗÌУ	Gly	 Met	•••				

Figure 11d (Continued)

AEFII la β-mannosidase (63GR)

5' ATG CTA CCA GAA GAG TTC CTA TGG GCC CTT CCC 245
The second secon
Met Leu Pro Glu Glu Phe Leu Trp Gly Val Gly Gln Ser Gly Phe Gln Phe Gli
63 72 21 Ser Gly Phe Gln Phe Gli
ATG GGC GAC AAG CTC AGG AGG CAC ATC GAT CCA AAT ACG GAD
Het Gly Asp Lys Leu Arg Arg His Ile Asp Pro Asn Thr Asp Trp Trp Lys Trp
117 126 135
GTT CGC GAT CCT TTC AAC ATA AAA AAG GAG CTT GTG AGT GGG GAC CTT CCC GAG Val Arg Asp Pro Pho Acc
Val Arg Asp Pro Phe Asn Ile Lys Lys Glu Leu Val Ser Gly Asp Leu Pro Glu
Lys Lys Glu Leu Val Ser Gly Asp Leu Pro Glu
GAC GGC ATC AAC AAC TAG GAA 189 198 207
GAC GGC ATC AAC AAC TAC GAA CTT TTT GAA AAC GAT CAC AAG CTC GCT AAA GGC ASD Gly Ile Asd ASD CTC GCT AAA GGC
Asp Gly Ile Asn Asn Tyr Glu Leu Phe Glu Asn Asp His Lys Leu Ala Lys Gly
225 234 243
CTT GGA CTC AAC GCA TAC AGG ATT GGA ATA GAG TGG AGC AGA ATC TTT CCC TGG
Leu Gly Leu Asn Ala Tyr Arg The Go
Leu Gly Leu Asn Ala Tyr Arg Ile Gly Ile Glu Trp Ser Arg Ile Phe Pro Trp
CCG ACG TGG ACG GTC GAT ACG GTC 315
CCG ACG TGG ACG GTC GAT ACC GAG GTC GAG TTC GAC ACT TAC GGT TTA GTA AAG
Pro Thr Trp Thr Val Asp Thr Glu Val Glu Phe Asp Thr Tyr Gly Leu Val Lys
333 342 351
CAC GTT AAG ATA GAC AAG TCC ACC CTT GCT GAA CTC GAC AGG CTG GCC AAC AAG
Asp Val Lys Ile Asp Lys Ser Thr Leu Ala Glu Leu Asp Arg Leu Ala Asn Lys
387 305
GAG GAG GTA ATG TAC TAC ACC COS 414 423
GAG GAG GTA ATG TAC TAC AGG CGC GTT ATT CAG CAT TTG AGG CAG CTC GGC TTC
Glu Glu Val Met Tyr Tyr Arg Arg Val Ile Gln His Leu Arg Glu Leu Gly Phe
441 450 459
AAG GTC TTC GTT AAC CTC AAC CAC TTC ACG CTT CCA ATA TGG CTC CAC GAC CCG
Lys Val Phe Val Asn Leu Asn Wil Di
Lys Val Phe Val Asn Leu Asn His Phe Thr Leu Pro Ile Trp Leu His Asp Pro
ATA GTG GCA AGG GAG AAG GCC GTG 533 522 531 540
ATA GTG GCA AGG GAG AAG GCC CTC ACA AAC GAC AGA ATC GGC TGG GTC TCC CAG
Ile Val Ala Arg Glu Lys Ala Leu Thr Asn Asp Arg Ile Gly Trp Val Ser Cla

Figure 120

							_						-· ·				
			AEP:	II :	La B	-mar	1208	ldas	• (630	B1)	(co	nti	aued)		
		54															
λG	G AC	A GT	r GT	T GA	C TT	T CC	- 44	יסכ יצידי ב	, c~		576	,		589	5		594 C GGJ
Ar	g Th	r Va.	l Va	ı Gı	u Ph	e Ala	Ly:	Tyr	. Ale	Ala	ı Tyr	Tle				·	
															, VT	ı Lei	1 617
GA	ר ה	60;		~ .~	61	2		621	•		630)		639	•		648
				- AL	A 1G	G AGC	. ACC	TIC	: AAC	: CY	r cci	, VIC	CI)	CIT	. CIC	GAC	648 CTC
Ası	Le	u Va	l As	Th	Tr	o Sei	Thr	. Phe									Leu
						, ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,			, v21	ı Gıı	Pro	Met	: Val	L Val	. Val	Gli	1 Leu
		657			666	5		675	;		684			697	i.		
GGC	TAC		: cc	- cc	TAC	: TC	CCX	TIT	. ccc	: ccc	GGA	GTC	ATC	AAC	. בכר	· cac	702 GCC
91)	171	. Let	T VIS	PY	TY:	: Ser	Gly	Phe	Pro	Pro	Gly	Val	Met	. Asn	Pro	Glu	Ala
	•	711	L		720	· ·		720									
GCG	; AAG	בת כ	GCC	ATO	CTC	. AAC	ATC	723 ATA	110	CCC	738			747			756 ATG
Ala	Lys	Leu	Ala	Ile	Leu	zeA ı	Met	Ile	λεη	Ala	His	λla	Leu	Ala	T	Tarm	Met
		765													•3-	בעם	net
מדג	330				774			783			792			801			810
					ACC	AAG	AAG	GCC	GAT	GAG	GAT	AGC	AAG	TCC	CCI	GCG	810 GAC
Ile	Lys	Arg	Phe	λsp	Thr	LVS	Lve	212	3 =	Clu	\						Asp
				_			_,_		ىرى.	310	ىردىم	SET	pys	Ser	Pro	λla	Asp
~~~		819			828			837			846			855			864
GPT	GGC	ATA	ATT	TAC	AAC	AAC	ATC	GGT	CLI	CCC	TAC	CCT	AAA	GAC	CCI	λλC	864 Gat
	3			-,3-	V211	ASI	TTG	GIY	Val	Ala	TAX	Pro	Lys	yzb	bio	λsn	Asp
		873			882		•	891			900			909			
CCC	AAG	GAC	GTŤ	AAA	GCA	GCC	GAA	AAC	GAC	AAC	TAC	TTC	CAC	303	CCN	~~~	918 TTC
PIO	rys	ASP	Val	Lys	Ala	Ala	Glu	Asn	λsp	yzu	Tyr	Phe	His	Ser	Gly	Leu	Phe
		927			936			945									
TTT	GAT	GCC	ATC	CAC	AAG	GGT	AAG	242	220	a Tra	954			963 GGC	_		972
Phe	Asp	λla	Ile	His	Lys	Gly	Lys	Leu	λsn	Ile	Glu	Phe	Asp	Glv	Glu	7	Phe
				•										0.1	914	نتحد	rne
GTA	222	981	) C	~~~	990			999		·ì	1008		3	L017		1	1026
			707		CTA	AAA	GGC	AAT	CYC	TGG	ATA	GGC	CIC	AAC	TAC	TAC	ACC
														Asn			
			-						بړمہ	1	***	отА	ren	Asn	Tyr	Tyr	Thr
		1035		. 1	1044		1	053		1	062		. 1	1071		1	.080
CGC	GAC	CLL	CIT	λGλ	TAT	TCG	GAG	CCC .	AAG	TTC	CCA	AGT	ATA	.071	~~	7	~~~

Figure 12b(Continued)

Arg Glu Val Val Arg Tyr Ser Glu Pro Lys Phe Pro Ser Ile Pro Leu Ile Ser

#### ABPII 1a $\beta$ -mannosidase (63GB1) (continued) 1098 TTC ANG GGC GTT CCC ANC THE GGC THE TCC TGC AGG CCC GGC ACG ACC TCC GCC --- --- --- --- --- --- --- --- --- --- --- --- ---Phe Lys Gly Val Pro Asn Tyr Gly Tyr Ser Cys Arg Pro Cly Thr Thr Ser Ala GAT GGC ATG CCC GTC AGC GAT ATC GGC TGG GAA GTC TAT CCC CAG GGA ATC TAC 1152 ---Asp Gly Met Pro Val Ser Asp Ile Gly Trp Glu Val Tyr Pro Gln Gly Ile Tyr GAC TCG ATA GTC GAG GCC ACC AAG TAC AGT GTT CCT GTT TAC GTC ACC GAG AAC --- --- --- --- --- --- --- --- --- --- --- --- ---Asp Ser Ile Val Glu Ala Thr Lys Tyr Ser Val Pro Val Tyr Val Thr Glu Asn 1251 1260 GGT GTT GCG GAT TCC GCG GAC ACG CTG AGG CCA TAC TAC ATA GTC AGC CAC GTC --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---Gly Val Ala Asp Ser Ala Asp Thr Leu Arg Pro Tyr Tyr Ile Val Ser His Val 1314 TCA AAG ATA GAG GAA GCC ATT GAG AAT GGA TAC CCC GTA AAA GGC TAC ATG TAC --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---Ser Lys Ile Glu Glu Ala Ile Glu Asn Gly Tyr Pro Val Lys Gly Tyr Met Tyr 1359 1368 TGG GCG CTT ACG GAT AAC TAC GAG TGG GCC CTC GGC TTC AGC ATG AGG TTT GGT Trp Ala Leu Thr Asp Asn Tyr Glu Trp Ala Leu Gly Phe Ser Met Arg Phe Gly 1413 1422 CTC TAC AAG GTC GAC CTC ATC TCC AAG GAG AGG ATC CCG AGG GAG AGA AGC GTT --- --- --- --- --- --- --- --- --- --- --- --- ---Leu Tyr Lys Val Asp Leu Ile Ser Lys Glu Arg Ile Pro Arg Glu Arg Ser Val 1467 1476 GAG ATA TAT CGC AGG ATA GTG CAG TCC AAC GGT GTT CCT AAG GAT ATC AAA GAG --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---Glu Ile Tyr Arg Arg Ile Val Gln Ser Asn Gly Val Pro Lys Asp Ile Lys Glu 1530 GAG TTC CTG AAG GGT GAG GAG AAA TGA 3' --- --- --- --- --- ---Glu Phe Leu Lys Gly Glu Glu Lys ***

Figure 12C(Continued)

#### OCI/4V Endoglucanase (33GP1)

9 18 27
5' ATG GTA GAA AGA CAC TTC AGA TAT GTT CTT ATT TGC ACC CTG TTT CTT GTT ATG
Met Val Glu Arg Nie De
Met Val Glu Arg His Phe Arg Tyr Val Leu Ile Cys Thr Leu Phe Leu Val Met
CTC CTA ATC TCA TCC ACT CAG TGT GGA ANA AAT GAA CCA AAC ANA AGA GTG AAT
Leu Leu Ile Ser Ser The Cla Co.
Leu Leu Ile Ser Ser Thr Gln Cys Gly Lys Asn Glu Pro Asn Lys Arg Val Asn
AGC ATG GAA GAG TO 126 135 144
AGC ATG GAA CAG TCA GTT GCT GAA AGT GAT AGC AAC TCA GCA TTT GAA TAC AAC
Ser Met Glu Gln Ser Val Ala Glu Ser Asp Ser Asn Ser Ala Phe Glu Tyr Asn
ASP Ser Ash Ser Ala Phe Glu Tyr Ash
AAA ATG GTA GGT AAA GGA GTA AAT ATT GGA AAT GCT TTA GAA GCT CCT TTC GAA
Lys Met Val Gly Lys Gly Val Asn Ile Gly Asn Ala Leu Glu Ala Pro Phe Glu
4/5 71/
GGA GCT TGG GGA GTA AGA ATT GAG GAT GAA TAT TTT GAG ATA ATA
GIV ALL THE GAR ATA ATA ANG ANA AGG
Gly Ala Trp Gly Val Arg Ile Glu Asp Glu Tyr Phe Glu Ile Ile Lys Lys Arg
4/9 200 -
GGA TTT GAT TCT GTT AGG ATT CCC ATA AGA TGG TCA GCA CAT ATA TCC GAA AAG
Gly Phe Asp Ser Val Are Yla
Gly Phe Asp Ser Val Arg Ile Pro Ile Arg Trp Ser Ala His Ile Ser Glu Lys
333 349
CCA CCA TAT GAT ATT GAC AGG AAT TIC CTC GAA AGA GTT AAC CAT GTT GTC GAT
Pro Pro Tyr Asp Ile Asp Arg Asn Phe Leu Glu Arg Val Asn His Val Val Asp
AGG GCT CTT GAG AAT AAT TTD AGG 405 414 423 437
AGG GCT CTT GAG AAT AAT TTA ACA GTA ATC ATC AAT ACG CAC CAT TTT GAA GAA
Arg Ala Leu Glu Asn Asn Leu Thr Val Ile Ile Asn Thr His His Phe Glu Glu
· GG! ACA
CTC TAT CAA GAA CCG GAT AAA TAC GGC GAT GTT TTG GTG GAA ATT TGG AGA CAG
Leu Tyr Gla Clu Dan
Leu Tyr Gln Glu Pro Asp Lys Tyr Gly Asp Val Leu Val Glu Ile Trp Arg Gln
977 604
ATT GCA AAA TTC TTT AAA GAT TAC CCG GAA AAT CTG TTC TTT GAA ATC TAC AAC
TAC AAC
Ile Ala Lys Phe Phe Lys len The
Ile Ala Lys Phe Phe Lys Asp Tyr Pro Glu Asn Leu Phe Phe Glu Ile Tyr Asn

Figure 13a

OC1/4V Endoglucanase (33GP1) (continued)
GAG CCT GCT CAG AAC TOTAL
SET CAG AAC TTG ACA GCT GAA AAA TGG AAC CG 585 594
GAG CCT GCT CAG AAC TTG ACA GCT GAA AAA TGG AAC GCA CTT TAT CCA AAA GTG Glu Pro Ala Gln Asn Leu Thr Ala Glu Lys Trp Asn Ala Leu Tyr Pro Lys Val
603 612 621
CTC AAA GTT ATC AGG GAG AGC AAT CCA ACC CGG ATT GTC ATT
Leu Lys Val Ile Arg Glu Ser Asn Pro Thr Arg Ile Val Ile Ile Asp Ala Pro
657 666 675 684
ARE THE GCA CAC TAT AGE GCA GTG AGA AGT CTA 122 693 702
AAC TGG GCA CAC TAT AGC GCA GTG AGA AGT CTA AAA TTA GTC AAC GAC AAA CGC Asn Trp Ala His Tyr Ser Ala Val Arg Ser Leu Lys Leu Val Asn Asp Lys Arg
ATC ATT GTT TCC TTC CAT TAC TAC GAA CCT TTC AND TOTAL 747
ATC ATT GTT TCC TTC CAT TAC TAC GAA CCT TTC AAA TTC ACA CAT CAG GGT GCC  Ile Ile Val Ser Phe His Tyr Tyr Glu Pro Phe Lys Phe Thr His Gln Gly Ala
765 774
GAN TOG GTT ANT CCC ATC CCA CCT GTT AGG GTT ANG TOG ANT GGC GAG GAN TOG
GIV THE WALL TOO GAR GAR TOO
And Pro lie Pro Pro Val Arg Val Lys Trp Asn Gly Glu Glu Trp
S19 S28 S37 S46 S55 S64  GAA ATT AAC CAA ATC AGA AGT CAT TTC AAA TAC GTG AGT GAC TGG GCA AAG CAA  G1u 11e Arr G1r T1r
Glu Ile Asn Gln Ile Arg Ser His Phe Lys Tyr Val Ser Asp Trp Ala Lys Gln
The Lys Tyr Val Ser Asp Trp Ala Lys Gln
AAT AAC GTA CCA ATC TIT CIT GGT GAA TTC GGT GCT TAT TCA AAA GCA GAC ATG
ASS GAN THE GGT GCT TAT TEX ANA GEN GNE ATG
Asn Asn Val Pro Ile Phe Leu Gly Glu Phe Gly Ala Tyr Ser Lys Ala Asp Het
GAC TCA AGG GTT AAG TGG ACC GAA AGT GTG AGA AAA ATG GCG GAA GAA TTT GGA
ASP Ser ATT VAL TOWN TO THE TOWN THE GOA
Asp Ser Arg Val Lys Trp Thr Glu Ser Val Arg Lys Met Ala Glu Glu Phe Gly
TOZO
Phe Ser Tyr Ala Tyr Trp Glu Phe Cys Ala Gly Phe Gly Ile Tyr Asp Arg Trp
TCT CAA AAC TOG ATC GAA CCA TTG GCA ACA GCT GTG GTT GGC ACA GGC AAA GAG Ser Glo Ard Tog Atc GAA CCA TTG GCA ACA GCT GTG GTT GGC ACA GGC AAA GAG
Ser Gln Asn Trp Ile Glu Pro Leu Ala Thr Ala Val Val
Ash Tip He Glu Pro Leu Ala Thr Ala Val Val Gly Thr Gly Lye Clu
TAA 3.
***

Figure 136(Continued)

#### Thermotoga maritima Fullulanase (60P3)

				9			1	0												
5.	λT(	G G	T	CII	ACJ	AA.	. GT		N.	·	27 .			3	6		4	5		ندد ه
												,16	AGC	CI(	S YY	CA	G TG	G CA	G GC	A AN
	Met	: As	p I	Leu	Thi	Ly	Va.	Gl	y I	le I	le V	al	λrg	Le	ı Ası	n G1	u Tr	p G1:		a Lys
				63			٦.	•					•							
		. 61	<b>G</b> (	CA.	AAA	GAC	: AGC	TT	C Y	A G	NG A	TA	λλλ	GAC	: GGJ		יב יצ	r cai	. ~	108 G TGG
	Asp	Va	1 )	la	Lvs	Acr														- 1GG
					_, _	,	, ,,,,,	· PA	E 11	e G	lu I	le	Lys	Asp	C13	Lys	Ala	Gl:	Va.	1 Trp
			1	.17			176				_									
	ATA	CI	c c	λG	GGA	GTG	GXA	GAO	3 AT	T 77	C T	AC	GAA	164		C) 6	15:	3		162 2 AGA
	T10	T														GAC	: AC	LTCI	, cc	AGA
	116	1-8	u G	īÞ	GIA	Val	Glu	Gli	ı Il	e Pb	e T	yΙ	Glu	Lys	Pro	Aso	The			· ···· · Arg
		·	1	71		٠	180											. Set	PIC	) Arg
	ATC	TIC	T	TC	GCA	CAG	GCA	AGG	- m-	18	9			198			207	•		216 AAT
												۹G	GIG	ATC	GAG	GCT	TIT	CIG	λCC	AAT
	Ile	Phe	2 P	þe	Ala	GTD	Ala	λrg	Se	r As	n Ly	73	Val	Ile	Glu	A7-				Asn
				25							_		-			~~4	PDG	Leu	Thr	ysu
	CCI.	OTO	. G	LJ AT	300	121	234			24	3			252			261			270
							746	AAA	GA	CI	C T	·C 2	AAG	GTT	ACT	GII	GAC	GGA	λλλ	270 GAG
	Pro	Val	. <b>λ</b> :	gp '	Thr	Lys	Lys	Lvs	Gla	ים! ום!										Glu
							-	•		- 20.	4 7		-ys	ATT	Thr	Val	ysb	GJA	Lys	Glu
	שיייר ע		27	9			288			297	7			306			315			•••
			G.		ICA.	YCX	GTG	GAA	λλG	GCC	: GA	T C	cc.	ACG	GAC	ATA	GAC	CAC	300	324 AAC
	Ile	Pro	٧a	1 9	Ser	Ara	 V=1											-:-		
								<b>314</b>	rys	ALE	l As	PE	LO.	Thr	yzd	Ile	yzb	Val	Thr	Asn ·
			33	3			347													
	TAC	GTG 	λG	λ)	TTC (	CIC	CIT	TCT	GAA	TCC	CI	Gλ	<b>A</b> A (	GAA	GAA	GAC	369	202		378
	IVE.	Val	Ar	 - T														vev	***	GAC
	JAT.			9 -	. 16	VAI .	ren	Ser	Glu	Ser	Le	a L	ys (	Glu	Glu	Asp	Leu	Arg	Lvs	λsp
			381	7		•	306													
(	TG (	SAA	CIY	3 A	TC )	ATA (	SAA (	<b>G</b> T	TAC	λλλ	ccc		נגם	res (	~~~		423			432
,	 /al (														·	AIC .	ATG	ATG	GAG	ATC
	/al (	<b>31</b> u	rei	7 T	Te 1	ile (	ilu (	Sly	Tyr	Lys	Pro	<b>)</b> A.	laj	rg /	Val :	Ile :	Met	Mer (		
			441	L		,	50													
C	TG C	AC	GAC	T	AC I	י נידגי	'AC (	:AT	GCA.	459	~~~		. 4	68			477			486
-	TG C			-									A G	CC C	STA :	TAT	ICI (	CCY (	SAG .	AAG
I	eu A	Sp.	λsp	T	yr T	ג בג	እ <u>ተ</u> ነ	sp (	Gly	Glu	Leu	G	y A	la v	7a) 1	·				
			495											T		YE .	er :	rro (	ilu .	Lys
A	CG A	TA ·	-17C		:X כ	ጉር ጥ 5	04 SC m	~~ -		513	_		5	22		:	531			540
-	CG A							(	.cc	CII.	TCT	AA	G T	cc c	K AT	vac d	o ote	TT C	TC :	TTC
T	h <del>r</del> I	le i	Phe	λι	g V	al T	rp S	er I	). 	 Val.	Ser	1.04								
									-			~3	<i>-</i> 1.	۷ د	aı I	ys \	/al I	Leu I	æu l	?he

Figure 14a

Thermotoga maritime Pullulanase (6GP3) (continued)
549 Fra
AAA AAC GGA GAA GAC ACA GAA CCG TAC CAG GTT GTG AAC ATG GAA TAC AAG GGA Lys Asn Gly Glu Asn The Gl
LVS ASD GIVE CAN CAR CAR GOA
The Glu Asp Thr Glu Pro Tyr Gln Val
Lys Asn Gly Glu Asp Thr Glu Pro Tyr Gln Val Val Asn Met Glu Tyr Lys Gly
AAC GCC GDD 753 612 621
AAC GGG GTC TGG GAA GCG GTT GTT GAA GGC GAT CTC GAC GGA GTG TTC TAC CTC  Asn Gly Val TTP Gly Ale Trace of the control of the c
ASD CAN USA
ory var Trp Glu Ala Val Val Glu Gly Arm Val
Asn Gly Val Trp Glu Ala Val Val Glu Gly Asp Leu Asp Gly Val Phe Tyr Leu
757 666 675 684
TAT CAG CTG GAA AAC TAC GGA AAG ATC AGA ACA ACC GTC GAT CCT TAT TCG AAA  Tyr Gln Leu Glu Asp Tog Gl
THE CITY OF THE THE PART OF THE TAR TO ANA
are din Leu Glu Asn Tyr Gly Lys Ile Arg The man
Tyr Gln Leu Glu Asn Tyr Gly Lys Ile Arg Thr Thr Val Asp Pro Tyr Ser Lys
711 720 729 730
GCG GTT TAC GCA AAC AAC CAA GAG AGC GCC GTT GTG AAT CTT GCC AGG ACA AAC
Ala Wal Daniel and And CTT GCC AGG ACA AAC
val Tyr Ala Asn Asn Gln Glu Ser Ala Val Val
Ala Val Tyr Ala Asn Asn Gln Glu Ser Ala Val Val Asn Leu Ala Arg Thr Asn
CCA GAA CCA CCA CCA CCA CCA CCA CCA CCA
CCA GAA GGA TGG GAA AAC GAC AGG GGA CCG AAA ATC GAA GGA TAC GAA GAC GCG Pro Glu Gly TTD Glu A
Pro Gly Cly The The Table 19 Con The GAA GAC GCG
old Gly Trp Glu Asn Asp Arg Gly Pro Lys Ile Gly Ct
Pro Glu Gly Trp Glu Asn Asp Arg Gly Pro Lys Ile Glu Gly Tyr Glu Asp Ala
ATA ATC TAT GAA ATA CAC ATA GCG GAC ATC ACA GGA CTC GAA AAC TCC GGG GTA
B64
The The Tyr Glu The Rice To
Ile Ile Tyr Glu Ile His Ile Ala Asp Ile Thr Gly Leu Glu Asn Ser Gly Val
873 882 en
AAA AAC AAA GGC CTC TAT CTC GGG CTC ACC GAA GAA AAC ACG AAA GGA CCG GGC
THE COLUMN STREET OF THE COLUM
Lys Asn Lys Gly Leu Tyr Leu Gly Leu Thr Glu Glu Asn Thr Lys Gly Pro Gly
Led Gly Leu Thr Glu Glu Asn Thr Lys Gly Per Gly
927 936
GGT GTG ACA ACA GGC CTT TCG CAC CTT GTG GAA CTC GGT GTT ACA CAC GTT CAT
CAL CIT GIG GAA CIC GGT GIT ACA CAC GITT CAN
Gly Val Thr Thr Gly Leu Ser Wig Louis
Gly Val Thr Thr Gly Leu Ser His Leu Val Glu Leu Gly Val Thr His Val His
981 990 999 1000
ATA CTT CCT TTC TTT GAT TTC TAC ACA GGC GAC GAA CTC GAT AAA GAT TTC GAG
TIO THE CAN GAT THE GAG
Leu Pro Phe Phe Asp Phe Tyr Thr Gly han Ch
Ile Leu Pro Phe Phe Asp Phe Tyr Thr Gly Asp Glu Leu Asp Lys Asp Phe Glu
1035 1044 1053 1052
AAG TAC TAC AAC TGG GGT TAC GAT CCT TAC CTG TTC ATG GTT CCG GAG GGC AGA
LVS TO THE ATE GIT CCG GAG GGC AGA
Lys Tyr Tyr Asn Trp Gly Tyr Asp Pro Tyr Leu Phe Het Val Pro Glu Gly Arg
The Met Val Pro Clu Glv Arg

Figure 14b(Continued)

1000
1089
1089 1098 1107 1116 1125 1134 TAC TCA ACC GAT CCC AAA AAC CCA CAC ACG AGA ATC AGA GAA GTC AAA GAA ATG
THE NEW ACC ACA ATC ACA CAN
Tyr Ser Thr Asp Pro Lys Asn Pro His Thr Arg Ile Arg Glu Val Lys Glu Met
GTC AAA GCC CTT CAC AAA CAC CCT and GTC 1170 1179 1100
Val Lys Ala Leu His Lys His Gly Ile Gly Val Ile Wat The Wat The Cort
Val Lys Ala Leu His Lys His Gly Ile Gly Val Ile Met Asp Met Val Phe Pro
1197 1206 1215 1224 1233
CAC ACC TAC GGT ATA GGC GAA CTC TCT GCG TTC GAT CAG ACG GTG CCG TAC TAC
His The Tyr Gly Ile Cive of
His Thr Tyr Gly Ile Gly Glu Leu Ser Ala Phe Asp Gln Thr Val Pro Tyr Tyr
1251 1260 1269 1278 1287 1296 TTC TAC AGA ATC GAC AAG ACA GGT GCC TAT TTG AAC GAA ACG GT GCC TAT TTG AAC GT GT GCC TAT TTG AAC GT GT GCC TAT TTG AAC GT
TTC TAC AGA ATC GAC AAG ACA GGT GCC TAT TTG AAC GAA AGC GGA TGT GGT AAC
Phe Tyr Arg Ile Asp Lys The Class
Phe Tyr Arg Ile Asp Lys Thr Gly Ala Tyr Leu Asn Glu Ser Gly Cys Gly Asn
1305 1314 1323 1332 1343
GTC ATC GCA AGC GAA AGA CCC ATG ATG AGA AAA TTC ATA GTC GAT ACC GTC ACC  Val Ile Ala Ser Club
Val Ile Ala Ser Glu Arg Pro Met Met Arg Lya Pho Ila Wal
1350 The 116 Val Asp Thr Val Thr
TAC TGG GTA AAG GAG TAT CAC ATA GAC GCA TTC 100 1395 1404
ATA GAC GGA TTC AGG TTC CAT CAG ATA
Tyr Trp Val Lys Glu Tyr His Ile Asp Gly Phe Arg Phe Asp Gln Met Gly Leu
ATC GAC AAA AAG ACA ATG CTC GAA GTC GAA ACA GTC GAA AC
The same and the s
Ile Asp Lys Lys Thr Het Leu Glu Val Glu Arg Ala Leu His Lys Ile Asp Pro
ACT ATC ATT CTC TAC GGC GAA CCG TGG GGT GGA TGG GGA GCA CCG ATC AGG TTT
Thr Ile Ile Leu Tyr Gly Cla Par
Thr Ile Ile Leu Tyr Gly Glu Pro Trp Gly Gly Trp Gly Ala Pro Ile Arg Phe
1521 1530 1539 1548 1557 1566
THE SEC ACA CAC GTG GCA GCT TTC AAC CAR COM
Gly Lys Ser Asp Val Ala Gly Thr His Val Ala
Gly Lys Ser Asp Val Ala Gly Thr His Val Ala Ala Phe Asn Asp Glu Phe Arg
1575 1584 1593 1602 1611 1620 GAC GCA ATA AGG GGT TCC GTG TTC ANG GGG 1500
AAC CCG AGC GTC AAG CCA TOTO COM
Asp Ala Ile Arg Gly Ser Val Phe Asn Pro Ser Val Lys Gly Phe Val Het Gly
Sel val Lys Gly Phe Val Het Gly

Figure 14C(Continued)

Thermotogs marietan
Thermotoga maritima Pullulanase (6GP3) (continued)
1629 1638 1647 1656 1665 1674
THE SAN AND GAN ACC AND ATC AND AGG GOT COTT COTT COTT
GGA TAC GGA AAG GAA ACC AAG ATC AAA AGG GGT GTT GTT GGA AGC ATA AAC TAC
Gly Tyr Gly Lys Glu Thr Lys Ile Lys Arg Gly Val Val Gly Ser Ile Asn Tyr
and the second s
1683 1692 1701
GAC GGA AAA CTC ATC AAA ACT TO 1710 1719
1728
GAC GGA AAA CTC ATC AAA AGT TTC GCC CTT GAT CCA GAA GAA ACT ATA AAC TAC
The Lys Ser Phe Ala Leu Asp Pro Clu Co
Asp Gly Lys Leu Ile Lys Ser Phe Ala Leu Asp Pro Glu Glu Thr Ile Asn Tyr
1737 1746 1755 1764 1773 1773
1773 1782
GCA GCG TGT CAC GAC AAC CAC ACA CTG TGG GAC AAG AAC TAC CTT GCC GCC AAA
Ala Ala Cys His Asp Asn His Thr Leu Tro Asp Line
Ala Ala Cys His Asp Asn His Thr Leu Trp Asp Lys Asn Tyr Leu Ala Ala Lys
1791 1800 1800
GCT GAT AAG AAA AAG GAA TGG ACC GAA GAA GAA CTG AAA AAC GCC CAG AAA CTG
1835
Ala Asp Ive
Ala Asp Lys Lys Glu Trp Thr Glu Glu Glu Leu Lys Asn Ala Gln Lys Leu
1845 1054
1845 1854 1863 1872 1993
Ala Gly Ala Ile Lou Lou Con Total CAR GCT GTT CCT TTC CAC GGA GGG CAG
Ala Gly Ala Ile Leu Leu Thr Ser Gln Gly Val Pro Phe Leu His Gly Gln
The Leu His Gly Gly Cln
CAC TTC TGC AGG ACG ACT TTC AAC GAC AAC TCC TAC AAC GCC CCT ATC TCG
THE
Asp Phe Cys Arg Thr Thr Asn Phe Asn Asp Asp Sec Cor Arc TCG
John Ser Tyr Asn Ala Pro Ile Ser
1980 COC 111 (12m m) A Co
TIC ATA CAC CON
Ile Asn Gly Phe Asp Tyr Glu Arg Lys Leu Gla Pho The
Ile Asn Gly Phe Asp Tyr Glu Arg Lys Leu Gln Phe Ile Asp Val Phe Asn Tyr
2007 2016 2025
CAC AAG GGT CTC ATA AAA CTC AGA AAA GAA CAC CCT GCT TTC AGG CTG AAA AAC
THE ANN GAA CAC CCT GCT THE ACC
His Lys Gly Leu Ile Lys Leu Arg Lys Glu His Pro Ala Phe Arg Leu Lys Asn
bys Leu Arg Lys Glu His Pro Ala Phe Arm I
2061 2070 2070
2061 2070 2079 GCT GAA GAG ATC ANA AND CO. 2079 2088 2097
AND CAC CTG GAA TIT CTC CCG CCC CCC CCC
GCT GAA GAG ATC AAA AAA CAC CTG GAA TTT CTC CCG GGC GGG AGA AGA ATA GTT Ala Glu Glu Ile Lys Lys High
Lys Lys His Leu Clu Phe Leu Pro Clu Clu
Ala Glu Glu Ile Lys Lys His Leu Glu Phe Leu Pro Gly Gly Arg Arg Ile Val
2115 2124 2133 2142 2151 2160
2151 2160
GCG TTC ATG CTT AAA GAC CAC GCA GGT GGT GAT CCC TGG AAA GAC ATC GTG GTG Ala Phe Met Law Law Cac GCA GGT GGT GAT CCC TGG AAA GAC ATC GTG GTG
Ala Phe Met Leu Lys Asp His Ala Gly Gly Asp Pro Trp Lys Asp Ile Val Val
TIP Lys Asp Ile Val Val

Figure 14d(Continued)

### Thermotoga maritima Pullulanasa (5093) (cont.)

	_												• -		~~~	.,	
ATT  Ile	Tyr J	(BU	Gly	λen	Leu	Glu	Lys	Thr	Thr	Tyr	Lys	Leu	Pro	 Glu	GGA  Glv	AAA  Tave	2214 TGG
<b>AAT</b>	CTG C	TT al	GTG  Val	AAC  Asn	AGC  Ser	CAG  Gln	777  Ly3	2241 GCC  Ala	GGA Gly	ACA Thr	2250 GAA  Glu	GTG  Val	ATA	2259 GAA  Glu	NÁO		2268
GGA	22 ACA A' : Thr I:	TA (	GAA		CAT	CCG	CLI.	2295 TCC	GCG	TAC	2304 GTT	CTG	TAC	2313 AGA	GAG	<b>~</b>	

Figure 14c(Continued)

Figure 15a Thermotoga maritima MSB8 (Clone # 6GP2) Glycosidase

CTT TTA TTG ATC GTT GAG CTC TCT TTC GTT CTC TTT GCA AGT GAC GAG TTC Leu Leu Leu Ile Val Glu Leu Ser Phe Val Leu Phe Ala Ser Asp Glu Phe

GTG AAA GTG GAA AAC GGA AAA TTC GCT CTG AAC GGA AAA GAA TTC AGA TTC Val Lys Val Glu Asn Gly Lys Phe Ala Leu Asn Gly Lys Glu Phe Arg Phe

ATT GGA AGC AAC TAC TAC ATG CAC TAC AAG AGC AAC GGA ATG ATA GAC Ile Gly Ser Asn Asn Tyr Tyr Met His Tyr Lys Ser Asn Gly Met Ile Asp

AGT GTT CTG GAG AGT GCC AGA GAC ATG GGT ATA AAG GTC CTC AGA ATC TGG Ser Val Leu Glu Ser Ala Arg Asp Met Gly Ile Lys Val Leu Arg Ile Trp

GGT TTC CTC GAC GGG GAG AGT TAC TGC AGA GAC AAG AAC ACC TAC ATG CAT Gly Phe Leu Asp Gly Glu Ser Tyr Cys Arg Asp Lys Asn Thr Tyr Met His

CCT GAG CCC GGT GTT TTC GGG GTG CCA GAA GGA ATA TCG AAC GCC CAG AGC Pro Glu Pro Gly Val Pne Gly Val Pro Glu Gly Ile Ser Asn Ala Gln Ser

-GGT TTC GAA AGA CTC GAC TAC ACA GTT GCG AAA GCG AAA GAA CTC GGT ATA Gly Phe Glu Arg Leu Asp Tyr Thr Val Ala Lys Ala Lys Glu Leu Gly Ile

AAA CTT GTC ATT GTT GTG AAC AAC TGG GAC GAC TTC GGT GGA ATG AAC Lys Leu Val lle Val Leu Val Asn Asn Trp Asp Asp Phe Gly Gly Met Asn

CAG TAC GTG AGG TGG TTT GGA GGA ACC CAT CAC GAC GAT TTC TAC AGA GAT Gln Tyr Val Arg Trp Phe Gly Gly Thr His His Asp Asp Phe Tyr Arg Asp

GAG AAG ATC AAA GAA GAG TAC AAA AAG TAC GTC TCC TTT CTC GTA AAC CAT Glu Lys Ile Lys Glu Glu Tyr Lys Lys Tyr Val Ser Phe Leu Val Asn His

GTC AAT ACC TAC ACG GGA GTT CCT TAC AGG GAA GAG CCC ACC ATC ATG GCC Val Asn Thr Tyr Thr Gly Val Pro Tyr Arg Glu Glu Pro Thr Ile Met Ala

TGG GAG CTT GCA AAC GAA CCG CGC TGT GAG ACG GAC AAA TCG GGG AAC ACG Trp Glu Leu Ala Asn Glu Pro Arg Cys Glu Thr Asp Lys Ser Gly Asn Thr

CTC GTT GAG TGG GTG AAG GAG ATG AGC TCC TAC ATA AAG AGT CTG GAT CCC Leu Val Glu Trp Val Lys Glu Met Ser Ser Tyr Ile Lys Ser Leu Asp Pro

AAC CAC CTC GTG GCT GTG GGG GAC GAA GGA TTC TTC AGC AAC TAC GAA GGA Asn His Leu Val Ala Val Gly Asp Glu Gly Phe Phe Ser Asn Tyr Glu Gly

TTC AAA CCT TAC GGT GGA GAA GCC GAG TGG GCC TAC AAC GGC TGG TCC GGT Phe Lys Pro Tyr Gly Glu Ala Glu Trp Ala Tyr Asn Gly Trp Ser Gly

GTT GAC TGG AAG AAG CTC CTT TCG ATA GAG ACG GTG GAC TTC GGC ACG TTC Val Asp Trp Lys Lys Leu Leu Ser Ile Glu Thr Val Asp Phe Gly Thr Phe

CAC CTC TAT CCG TCC CAC TGG GGT GTC AGT CCA GAG AAC TAT GCC CAG TGG His Leu Tyr Pro Ser His Trp Gly Val Ser Pro Glu Asn Tyr Ala Gln Trp

GGA GCG AAG TGG ATA GAA GAC CAC ATA AAG ATC GCA AAA GAG ATC GGA AAA Gly Ala Lys Trp Ile Glu Asp His Ile Lys Ile Ala Lys Glu Ile Gly Lys

CCC GTT GTT CTG GAA GAA TAT GGA ATT CCA AAG AGT GCG CCA GTT AAC AGA Pro Val Val Leu Glu Glu Tyr Gly Ile Pro Lys Ser Ala Pro Val Asn Arg

ACG GCC ATC TAC AGA CTC TGG AAC GAT CTG GTC TAC GAT CTC GGT GGA GAT Thr Ala Ile Tyr Arg Leu Trp Asn Asp Leu Val Tyr Asp Leu Gly Gly Asp

GGA GCG ATG TTC TGG ATG CTC GCG GGA ATC GGG GAA GGT TCG GAC AGA GAC Gly Ala Met Phe Trp Met Leu Ala Gly Ile Gly Glu Gly Ser Asp Arg Asp

GAG AGA GGG TAC TAT CCG GAC TAC GAC GGT TTC AGA ATA GTG AAC GAC GAC Glu Arg Gly Tyr Tyr Pro Asp Tyr Asp Gly Phe Arg Ile Val Asn Asp Asp

AGT CCA GAA GCG GAA CTG ATA AGA GAA TAC GCG AAG CTG TTC AAC ACA GGT Ser Pro Glu Ala Glu Leu Ile Arg Glu Tyr Ala Lys Leu Phe Asn Thr Gly

GAA GAC ATA AGA GAA GAC ACC TGC TCT TTC ATC CTT CCA AAA GAC GGC ATG Glu Asp Ile Arg Glu Asp Thr Cys Ser Phe Ile Leu Pro Lys Asp Gly Met

GAG ATC AAA AAG ACC GTG GAA GTG AGG GCT GGT GTT TTC GAC TAC AGC AAC

Figure 15b (continued)

Glu Ile Lys Lys Thr Val Glu Val Arg Ala Gly Val Phe Asp Tyr Ser Asn

ACG TTT GAA AAG TTG TCT GTC AAA GTC GAA GAT CTG GTT TTT GAA AAT GAG Thr Phe Glu Lys Leu Ser Val Lys Val Glu Asp Leu Val Phe Glu Asn Glu

ATA GAG CAT CTC GGA TAC GGA ATT TAC GGC TTT GAT CTC GAC ACA ACC CGG Ile Glu His Leu Gly Tyr Gly Ile Tyr Gly Phe Asp Leu Asp Thr Thr Arg

ATC CCG GAT GGA GAA CAT GAA ATG TTC CTT GAA GGC CAC TTT CAG GGA AAA Ile Pro Asp Gly Glu His Glu Met Phe Leu Glu Gly His Phe Gln Gly Lys

ACG GTG AAA GAC TCT ATC AAA GCG AAA GTG GTG AAC GAA GCA CGG TAC GTG Thr Val Lys Asp Ser Ile Lys Ala Lys Val Val Asn Glu Ala Arg Tyr Val

CTC GCA GAG GAA GTT GAT TTT TCC TCT CCA GAA GAG GTG AAA AAC TGG TGG Leu Ala Glu Glu Val Asp Phe Ser Ser Pro Glu Glu Val Lys Asn Trp Trp

AAC AGC GGA ACC TGG CAG GCA GAG TTC GGG TCA CCT GAC ATT GAA TGG AAC Asn Ser Gly Thr Trp Gln Ala Glu Phe Gly Ser Pro Asp Ile Glu Trp Asn

GGT GAG GTG GGA AAT GGA GCA CTG CAG CTG AAC GTG AAA CTG CCC GGA AAG Gly Glu Val Gly Asn Gly Ala Leu Gln Leu Asn Val Lys Leu Pro Gly Lys

AGC GAC TGG GAA GAA GTG AGA GTA GCA AGG AAG TTC GAA AGA CTC TCA GAA Ser Asp Trp Glu Glu Val Arg Val Ala Arg Lys Phe Glu Arg Leu Ser Glu

TGT GAG ATC CTC GAG TAC GAC ATC TAC ATT CCA AAC GTC GAG GGA CTC AAG Cys Glu Ile Leu Glu Tyr Asp Ile Tyr Ile Pro Asn Val Glu Gly Leu Lys

GGA AGG TTG AGG CCG TAC GCG GTT CTG AAC CCC GGC TGG GTG AAG ATA GGC Gly Arg Leu Arg Pro Tyr Ala Val Leu Asn Pro Gly Trp Val Lys Ile Gly

CTC GAC ATG AAC AAC GCG AAC GTG GAA AGT GCG GAG ATC ATC ACT TTC GGC Leu Asp Met Asn Asn Ala Asn Val Glu Ser Ala Glu Ile Ile Thr Phe Gly

GGA AAA GAG TAC AGA AGA TTC CAT GTA AGA ATT GAG TTC GAC AGA ACA GCG Gly Lys Glu Tyr Arg Arg Phe His Val Arg Ile Glu Phe Asp Arg Thr Ala

Figure 15C(continued)

GGG GTG AAA GAA CTT CAC ATA GGA GTT GTC GGT GAT CAT CTG AGG TAC GAT Gly Val Lys Glu Leu His Ile Gly Val Val Gly Asp His Leu Arg Tyr Asp

GGA CCG ATT TTC ATC GAT AAT GTG AGA CTT TAT AAA AGA ACA GGA GGT ATG Gly Pro Ile Phe Ile Asp Asn Val Arg Leu Tyr Lys Arg Thr Gly Gly Met

TGA

1991

END

Figure 15d(continued)

# Figure No. 16/ Thermotoga maritima MSB8(6gb4)

1 ATG AAA AGA ATC GAC CTG AAT CCT	
1 ATG AAA AGA ATC GAC CTG AAT GGT TTC TGG AGC GTT AGG GAT AAC GAA GGG AGA TTT TCG	
1 Met Lys Arg Ile Asp Leu Asn Gly Phe Trp Ser Val Arg Asp Asn Glu Gly Arg Phe Ser	60
61 TTT Can occurrence	- 20
THE GAR GGG ACT GTG CCA COG	
Phe Glu Gly Thr Val Pro Gly Val Val Gla Na CTG GTC AGA AAA GGT CTT CCA	120
21 Phe Glu Gly Thr Val Pro Gly Val Val Gln Ala Asp Leu Val Arg Lys Gly Leu Leu Pro	
121 CAC CCG TAC GTT GCG AMO AND	40
121 CAC CCG TAC GTT GGG ATG AAC GAA GAT CTC TTC AAG GAA ATA GAA GAC AGA GAG TGG ATC	
41 His Pro Tyr Val Gly Met Asn Glu Asp Leu Phe Lys Glu Ile Glu Asp Arg Glu Trp Ile	180
181 mag Arg Glu Trp Ile	60
181 TAC GAG AGG GAG TTC GAG TTC AAA GAA GAT GTG AAA GAG GGG GAA CGT GTC GAT CTC GTT 61 Tyr Glu Arg Glu Phe Glu Phe Lys Glu Asp Val Lys Glu Gy	
61 Tyr Glu Arg Glu Phe Glu Phe Lys Glu Asp Val	240
61 Tyr Glu Arg Glu Phe Glu Phe Lys Glu Asp Val Lys Glu Gly Glu Arg Val Asp Leu Val	80
241 TTT GAG GGC GTC GAC ACC COO.	
241 TTT GAG GGC GTC GAC ACG CTG TCG GAT GTT TAT CTG AAC GGT GTT TAC CTT GGA AGC ACC 81 Phe Glu Gly Val Asp Thr Leu Ser Asp Val TVT Leu Acc GCT GTT TAC CTT GGA AGC ACC	
81 Phe Glu Gly Val Asp Thr Leu Ser Asp Val Tyr Leu Asn Gly Val Tyr Leu Gly Ser Thr	300
301 GAA GAC ATG TOO ATG	100
OAC AIG TTC ATC GAG TAT COO TO	
101 Glu Asp Met Phe Ile Glu Tyr Arg Phe Asp Val Thr Asn Val Leu Lys Glu Lys Asn His	350
Ash val Leu Lys Glu Lys Ash His	120
361 CTG AAG GTG TAC ATA AAA TCT GGG ADD	
361 CTG AAG GTG TAC ATA AAA TCT CCC ATC AGA GTT CCG AAA ACT CTC GAG CAG AAC TAC GGG	. 430
121 Leu Lys Val Tyr Ile Lys Ser Pro Ile Arg Val Pro Lys Thr Leu Glu Gln Asn Tyr Gly	420
421 GTC CTC GCG GGT	140
421 GTC CTC GGC GGT CCT GAA GAT CCC ATC AGA GGA TAC ATA AGA AAA GCC CAG TAT TCG TAC	
141 Val Leu Gly Gly Pro Glu Asp Pro Ile Arg Gly Tyr Ile Arg Lys Ala Gln Tyr Ser Tyr	480
483 are	160
481 GGA TGG GAC TGG GGT GCC AGA ATC GTT ACA AGC GGT ATT TGG AAA CCC GTC TAC CTC GAG	
Gly Trp Asp Trp Gly Ala Arg Ile Val Thr Sor Cluster	540
161 Gly Trp Asp Trp Gly Ala Arg Ile Val Thr Ser Gly Ile Trp Lys Pro Val Tyr Leu Glu	180
541 GTG TAC AGG GCA CGT CTT CAG GAR	
541 GTG TAC AGG GCA CGT CTT CAG GAT TCA ACG GCT TAT CTG TTG GAA CTT GAG GGG AAA GAT 181 Val Tyr Arg Ala Arg Leu Gln Asp Ser Thr Ala Tyr Lev Ly	
181 Val Tyr Arg Ala Arg Leu Gln Asp Ser Thr Ala Tyr Leu Leu Glu Leu Glu Gly Lys Asp	600
601 GCC CTT CTT CTT CTT CTT CTT CTT CTT CTT	200
CIT GIG AGG GTG AAC CGT TTG GT	
201 Ala Leu Val Arg Val Asn Gly Phe Val His Gly Glu Gly Asn Leu Ile Val Glu Val Tyr	660
The Gry Ash Leu Ile Val Glu Val Tyr	220
661 GTA AAC GGT GAA AAG ATA GGG GAG	
Val Asn Gly Glu Lys Ile Gly Clu Di	720
Lys Ash Gly Clusters	720 240
721 GAT GGA GTG TTG CLA	240
721 GAT GGA GTG TTC CAC CTG AAA GAT GTG AAA CTA TGG TAT CCG TGG AAC GTG GGG AAA CCG	
241 Asp Gly Val Phe His Leu Lys Asp Val Lys Leu Trp Tyr Pro Trp Asn Val Gly Lys Pro	780
Try Asn Val Gly Lys Pro	260

	781	TAC	CTG	TAC	GAT	TTC	יירט :	الملك ا	· ~		<b>.</b>														
3	61	Tyr	Leu	TAC Tyr	Asp	Phe	Val	Dha	- G1	G 17	G AA	LA GA	C T	TA A	AC	GGA	GAG	ATO	TA	C AG	A G	AA C	AA	84	_
				Tyr	•	- 120		Pile	· va	r re	u Ly	's As	p Le	eu A	sn (	Gly	Glu	Πle	Ty:	r Ar	g G1	iu G	111		
																		•						28	U
	81 1		AAA	ATC Ile	GGT	TTG	AGA	AGA	GT	C AG	A AT	C GT	т са	AG G	AG (	200	GAT	GD 2	<b>C</b> 2.						
*	01 1	Jys	Lys	Ile	Gly	Leu	Arg	Arg	Va]	Ar	g I1	e Va	1 G1	n G	lu r	250	D.C.	C)	GAZ	. GG	а да	A A	CT.	906	3
																								300	)
9	01 7	TC I	ATA	TTC Phe	GAA	ATC	AAC	GGT	GAG														•		
3	01 P	he.:	Ile	Phe	Glu	Ile	Asn	Glv	Glu			C III	C GC	T A	AG G	GT	GCT	AAC	TGG	ATT	cc	C T	CA ·	960	)
				•				,	514	Lys	va.	I Pho	e Al	a Ly	/S G	ly.	Ala	Asn	Trp	Ile	Pr	0 St	er	320	
9																									
	21 G	]:: 2	SED :	ATC Ile :	C1C	ACG	TGG	TTG	AAG	GAG	GAZ	A GAT	TA	C GA	A A	AG (	CTC	GTC	AAA	ATC					
			.SII .	Ile :	Leu	Thr	Trp	Leu	Lys	Glu	Glu	a Asp	Ту	F G1	u L	ys 1	Leu	Ţ Val	Lare	Man		\ AG	G	1020	
																								340	
102	1 A	GT G	CC 3	LAT ,	ATG .	AAC	ATG	CTC	AGG	GTC	TGG	GGA	GGZ												
34	1 S	er A	la A	lsn M	let .	Asn	Met	Leu	Ara	Val	T	Gly	- C1.		A A	LC 1	'AC	GAG	AGA	GAG	ATO	TT	C	1080	
												Gly	GIY	/ G1;	у 1.	le 1	yr (	Glu	Arg	Glu	Ile	Ph	e.	360	
108	1 TA	C A	GA C	TC 1	GT (	7A.T	C 2 2																		
36	1 Ty	r A:	rq L	TC 1	`vs 1	250	CIM	CIC	GGT	ATC	ATG	GTG	TGG	CAC	G GA	T T	TC I	ATG	TAC	GCG	TGT	CT'	т	1140	
			_	eu C	., .	.sp	GIU.	ren	Gly	Ile	Met	Val	Trp	Glr	a As	p P	he N	let :	Tyr	Ala	Cvs	Lei		380	
114																								200	
38:	. GA		AT C	CG G	AT C	AT (	CTT (	CCG :	rgg	TTC	AGA	AAA	CTC	GCG	. AA	C G	א א כ	: A.C. 4						•	
50.	. 61	ս 13	YT P	ro A	sp H	lis 1	Leu	Pro :	rp	Phe	Arg	Lys	Leu	Ala	As	n G	lu c	2	SCA.	AGA	AAG -	ATI	Γ :	1200	
																								400	
1201	GT	G AG	A A	AA C' /S L	TC A	GA 1	CAC C	AT C	cc :	TCC	ል <del>ጉ</del> ጉ	CTT	CTC.			_									
401	. Va	l Ar	g Ly	/S L	eu A	rg I	yr H	lis F	ro s	Ser	Tle	V21	Tan	TGG	TG	C G	EA A	AC ;	40 (	GAA .	AAC	AAC	. 1	260	
				/S L								VAI	Leu	Trp	Cy:	s Gl	уА	sn A	sn (	Glu /	Asn	Asn		420	
1261																									
421	Trp	G1;	y Ph	C G	:D G	) .v. 1	c	GA A	AT A	ATG (	SCC .	AGA	AAA	GTG	GA:	r GG	T A	IC A	AC C	erc o	GA	AAC	1	320	
			-	e As			-p G	ту А	sn M	let )	Ala .	Arg	Lys	Val	Asp	G1	у 1	le A	sn L	eu (	ly.	Asn		440	
1321																									
441	Aro	T.o.	- IA	C CT	C TT	C G	AT T	TT C	CT G	AG A	TT :	TGT (	GCC	GAA	GAA	GA	c c	G T	רר א	CT 6		~ » ~			
	9	Det	ı ıy	r Le	u Ph	e A	sp P	he P	ro G	lu I	le (	ys )	Ala	Glu	Glu	As	D Pr	0 8	- T	h		TAT		380	
																							•	460	
1381	TGG	CCA	TC	C AG	T CC	A T	AC GO	SC GC	ST G	AA A	AA c	ברנה ז	, AC	N'CC	<i>~</i> >>										
461	Trp	Pro	Se	r Se	r Pr	0 T)	r G	y G	V G	lu L	ve A	ת בו		7	GAA	AA(	3 GA	A G	GA G	AC A	GG (	CAC	14	40	
									•	<b>-</b>	,		2511 2	ser	GIU	Lys	G1	u G	Ly A	sp A	rg H	iis	4	180	
1441	GTC	TGG	TAC	GT	TG(	G 20	T CC												٠,						
481	GTC Val	Trp	Tyr	Val	Tres	n Sa	- 0	- 16	G A	rg A	AC I	AC G	AA A	YAC .	TAC	GAA	AA.	A GA	C AC	C G	GA A	GG	1.5	00	
	Val	-				y se	r GI	y Tr	P Me	et As	sn T	yr G	lu A	sn :	Tyr	Glu	Ly:	s As	p T	r G	Lv A	ro		00	
1501																							,		
501	TTC Phe	AIC TI-	AGC	GAG	TT	r GG.	A TT	T CA	G GG	T GC	T C	cc c	AT C	CA C	SAC	ACC	b ጥ ፣		c	·					
	Phe	тте	ser	Glu	Phe	Gl	y Ph	e Gl	n Gl	y Al	a P	-о н	is p	ron	3311	Th-	71	. CA	G TT	C T7	TT	CA	15		
																							5:	20	
1561	AAA Lys	ccc	GAG	GAA	AGA	GAC	G AT	A TT	_ CA	T CC	'C		no -												
521	Lys	Pro	Glu	Glu	Arg	Gli	ı Ile	Phe	. Hi		. v.	. A.	יי בי	TG A	LAG	CAC	AAC	. AA	A CA	G GI	G G	٩A	163	20	
	Lys				_						o va	T ME	e Le	eu L	ys	His	Asn	Ly	s G1	n Va	1 G	lи	54	10	
									Fig	ure	16b	(cor	ntin	wed	)										

1623 543	_	A C	AG (	AA /	AGA	TT	G AT	C AC	G T	C A	ra i	rc Go	SA AJ	AT T	TT G	a a:	<b>A</b> C:	سن			-	C GA	
	٠.	. J . G.	-11	aru )	ug	Lei	1 I I	e Ar	g Pi	e I	le Pi	ie Gi	y As	n Pl	ie Gi	A		101	AA.	GA!	r TT	C GA(	1680
1681	L AG	T 7-1	· ·	·TO =												., .	ys	Cys	Lys	Ası	Ph	e Asp	560
561	. Se	r Ph	ie V	al T	yr	Leu	Se:	C CA	G CT n Le	C AA u As	n Gl	G GC n Al	G GA a Gl	G GC	G AT	C A	AG YS	TTC Phe	GGT Gly	GTT Val	GAJ	A CAC	1740
1741	TG	G CG	A A	GC A	GG	AAG	TAC	* AA:	h h ~														580
581	Tr	Ar _i	g Se	er A	rg	Lys	Tyr	Ly	Th	r Ala	a Gl	y Ala	Le	u Ph	C TG	G C2 P G1	G:	TTC Phe	AAC Asn	GAC Asp	AGC Ser	TGG	
1801	CCG	GT	77	C A	3C :	TGG	TCC	CC	CTC														600
601	Pro	Va]	l Ph	ie Se	er 1	Trp	Ser	Ala	· Val	Asp	Туг	Phe	Lys	AGG Arg	Pro	Ly.	A G S A	CT	CTC Leu	TAC Tvr	TAC	TAT	1860
1861	GCG	AGA	AG	A TI	C 1	TTC	GCT	GAA															620
621	Ala	Arg	Ar	g Ph	e P	he	Ala	Glu	Val	Leu	Pro	Val	Leu	Lys	AAG Lys	AG/ Arc	A G	AC A	AAC .	AAA	ATA	GAA	1920
1921	CTG	CTG	GT	G GG	ΤG	AG	CGA	TOT	C) C						-	•	,			Lys	116	Glu	640
641	Leu	Leu	Va.	[ G1	y G	lu .	Arg	Ser	Glu	GGA Gly	GAC Asp	AAA Lys	AGA Ara	AGT Ser	CTC	TCT	CZ	AG G	CT 7	rgc .	AGC	CTA	1980
L981	CGA	GAA	GAR	, cc		~» ·						•	-5			SeI	GI	n A	la (	ys :	Ser :	Leu	660
661	CGA Arg	Glu	Glu	Gly	A	rg 1	ys (	GGT Gly	ATT Ile	CGA Ara	AAA Lvs	GAC	TTA	CAG	AAC	GGT	AC	TC	CC A	GC #	IGA (	CGG	2040
								-		3	-, 0	vah	กรบ	GIN	Asn .	Gly	Th	r P	ro S	er A	irg #	Arg	680
041 681	Cys	Glu	TIT Phe	GGT Gly	En	A ad	205 685	_					٠.										

Figure 16¢(continued)

# Figure No. 17 Bankia gouldi (37gp4)

	1 4	TG	AAA	444	B B	- ~					•															
	i M	le+ 1	326	***	701	.ı c.	ra c	TA A	TG I	TT /	AAA	AG(	3 (7	T A	CG 1	ΤĄΤ	CT	A C	T	TG	TTI	TT	מ בי	TG	CTĠ	
			-ys	ьys	AS	n Le	eu L	eu M	et P	he 1	Lys	Arg	J Le	u T	hr 1	уr	Lei	u Pr	o I	eu	Phe	1.0	11 M	۰-۰	ton	60
6	1 0	TC 7	CA	CTA	AG	T TC	A G	TA G	CT C	AA 1	CT	CCI	GT	A C	ת ממ	B B	~									
2	1 L	eu S	er	Leu	Se	r Se	r V	al A	la G	ln s	Ser	Pro	Va	c	1 Y	~~	CA	r GG	ic c	GT	TTA	CA	A G	TT	GAC	120
				•	٠.								, vu		ru L	ys	HIE	G1	уА	rg	Leu	Gl	n V	<b>a</b> 1	Asp	40
12	1 G	GA A	AC (	CGC	AT"	ייט ז	י מידי	T 00	~ -											•						
4	1 G	ly A	sn i	Ara	114	 - T.e		AT GO	.6 11	JT G	GA (	GAA	AT	T A	CG A	GC	TTA	GC	T G	GT .	AAC	AG	: c:	rc	TTT	180
		•		3			u A:	[ג מו	.a S	er G	ly (	Glu	Il	e Ti	r s	er	Leu	Al	a G	ly i	Asn	Sei	. Le	eu.	Phe	60
18:																										
		A	GT J	AT	GCT	r GG	A GA	C AC	C T	C G	AT :	TT	TA:	r aa	T G	CA (	GAA	AC	T G'	רד (	SAT	لمىلى <u>.</u>	•			
0.	L T	p s	er )	lsn	Ala	Gl	y As	p Th	x Se	r A	sp 1	Phe	Ty	As	n A	la (	Glu	Th	r Va	1 1	aen 	221		А.	GCA	240
																										80
- 241	. G.	A A	AC 1	GG:	AAT	AG	C TC	A CT	T AI	T A	GA A	ATA	GCT	י א	e e		~~.			_						
81	G1	u A	sn I	rp .	Asn	Se	r Se	r Le	u Il	e A	ra I	le	Ala	Ma	- G		JIA	AA/	A G≱	A J	LAT	TGG	GA	T	GGC	300
											-5 -					.y \	Val	Lys	5 G]	u A	lsn	Tip	As	p	Gly	100
301	GG	A AJ	AT G	GC '	TAT	ATT	r (12	~ h~	T 00																	
101	G1	y As	n G	lv '	Tvr	Tle	. Dr	T AG	- 5-	G (;)	AG G	AG	CAA	GA	A GC	Τ >	\AA	ATI	AG	A A	AA	GTT	AT	T (	GAT	360
				•	- , -			p Se	L PI	0 G1	in G	lu	Gln	Gl	u Al	a I	γys	Ile	Ar	g L	ys	Val	11	e )	Asp	- 120
361																										
121	ומ	ת טנ ה או	. A	11 (	CT	AAC	: GG	C AT	TA	T GI	'A A'	TA	ATA	GA	TG	GC	AC	ACT	CA	C G	AA (	GCA	GAC	: า	מידים	420
	~	a WI	a 1.	Te 1	\la	Asn	Gl	/ Ile	Ty	r Va	1 1	le	lle	Ası	Tr	рн	is	Thr	Hi	s G	lu i	Ala	Gli	, 1	.011	140
																										140
421	TA	C AC	A G	AT G	AG	GCT	GT:	GAC	TT	TT	T A	cc .	AGA	ATG	: GC:	c	20	~~ »								·
141	Ty	Th	r As	g q	lu	Ala	Va]	Asp	Phe	: Ph	e Ti	ar .	Ara	Mer	. מ	n. G.	AC	C 1 A	TAG	. G(	GA (	AT	ACI	. с	CC.	480
													3			• ^.	sp	reu	тy	r G.	ly A	lsp	Thr	P	ro	160
481	AAT	GT.	A A	GI	ΆΤ	GAA	ATT	' ТАТ	אאר	. CD	~ ~~	<b></b> .			_			,								•
161	Asr	Va:	l Me	t T	yr	Glu	Ile	TAT	200	Cl			ATA	TAC	CAZ	A A	GT '	TGG	CCI	. Gi	TT A	TT	AAG	A.	AT	540
								Tyr	No.	GI	u PI	. 0	Lie	Tyr	Glr	Se	er :	Trp	Pro	Va	al I	le	Lys	A	sn	180
541	TAI	GC	מס ג		n n	CT1			_																	
181	Tyr	Ala	ເເດ		~~ 1 ~ ·	U-1	ATT	GCT	GGT	ATA	A CG	T	CT	AAA	GAC	. cc	-A (	GAT	AAT	TI.	'A A	TA.	ATT	G'	TA .	600
	•				111	val	116	Ala	Gly	Ile	Ar	g S	er	Lys	Asp	Pr	:0 }	Asp	Asn	Le	u I	le :	Ile	Va	al	200
601																										
201	GGT	ACI	AG	CA	AT '	TAT	TCT	CAG	CAA	GTI	GA	T G	TA	GCA	TCA	GC	:A 0	AC	CCA	ΔТ	ית מי	~~ <i>,</i>	~ ~ ~			
201	GIY	Thr	Se	r A	sn '	Tyr	Ser	Gln	Gln	Val	As	pV	al.	Ala	Ser	Al	аÀ	SD	Pro	71	~ c	'	3A1	A(	-1	660
																										220
661	AAT	GTG	GC	A TA	T J	ACT	TTA	CAT	TTT	TAT	פרי	ם מ	ري ر	معامل			_									
221	Asn	Val	Ala	а ту	'F ]	Chr	Leu	His	Phe	Tvr	יות:			a III	MAC	-	G C	TA	GAT	AA	C T	ra a	\GA	AA	T	720
								_		- , -	***	- ^	-a )	riie	ASD	Pr	о н	is.	Asp	As	n Le	eu A	urg	As	n	240
721	GTA	GCA	CAC	. אר	י מי		TT:																			
241	Val	Ala	G1 r	7 Th	, rs (	1-	ATA Las	GAT Asp	AAT -	AAT	GTT	G	CT 1	TG	TTT	GT:	T A	CA (	GAA	TG	G GC	T A	CA	AT	T	780
				• • • •	- 4	ud :	ren	Asp	Asn	Asn	Val	Α.	la I	eu	Phe	Va:	1 T	hr (	Glu	Tr	o G1	v T	hr	11	e	260

	781	TTA	AAT	ACC	GGA	CAA	CCN															
•	261	Leu	Asn	Thr	C1		GGA	GAA	CCA	GAC	AAA:	GAA	AGC	ACT	AAT	ACT	TGG	D. T.C.		TTT T		
•				****	GIY	GIN	Gly	Glu	Pro	Asp	Lys	Glu	Ser	Thr	Acn	The		NIG.	CCC	TTT T	TG 84	0
															ASII	inr	Trp	Met	Ala	Phe L	eu 28(	0
	841	AAA	GAA	AAA	GGT	ATA	ACT	~~~														
-	281	Lys	Glu	Lve	G) v	73-	WO1	CAC	GCT	AAT	TGG	TCT	TTG	AGT	GAC	AAA	GCT	ملحلمل		GAA A		
		-		-, -	GLY	TIE	Ser	His	Ala	Asn	Trp	Ser	Leu	Ser	Asp	Lare	212	*11	CCI	GAA AG Glu Th	CA 900	)
															,,,p	Dys	WIS .	Phe	Pro	Glu Th	ur 300	ļ
	901	GGG	TCT	GTA	GTT	CAA	GCA	GC »	C > >													
	301	Gly	Ser	Val	Val	Gla	21.	00A	CAA	GGT	GTA	TCT	GGT	TTA	ATT	AGC	AAT J	AAA	د بست	ACA GO		
						<b></b>	A1d	GIY	Gln	Gly	Val	Ser	Gly	Leu	Ile	Ser .	Asn t	310 1		ACA GC	C 960	
																		Jys 1	seu 1	Inr Al	a 320	
	961	TCT (	GGT	GAA :	ATT (	GTA :	AAA ;	AAC .	ATC	חדת	Chh									GA CC		
	321	Ser (	Gly (	Glu :	Ile v	Val 1	Lvs i	Acn .	T 7 -		CAA.	AAC	TGG (	GAT 2	ACA (	GAG 1	ACC I	CT A	CA G	GA CC	т 1000	
							-,-,	1011	iie .	TTE	Gln .	Asn :	Ltb 1	Asp :	Thr (	Glu 2	hr s	er T	י אמי	1	T 1020	
. 11	021																			TA PE	340	
	021	AAA A	ICA 3	ACA C	LAA T	GT A	AGT A	CT ;	ATA C	GAA '	TCT :	د مصدر								AA GCZ		
2	341	Lys .I	hr 1	hr c	iln C	ys s	er T	'hr 1	110.0				IGA G	CT C	CA A	TG G	AA A	CA G	CA C	AA GCZ	1080	
										31U (	Lys ]	le A	rg A	la A	la M	let G	lu T	hr A	la G	AA GCA ln Ala	350	
10	81 (	. 43 <del>5</del>	ስጥ ~																		360	
			MI G	AA A	TT A	TA A	TT G	cc c	CT G	GA A	AC I	AC A	אר דע	TT ~						T GCC		
	O	JY A	sp G	lu I	le I	le I	le A	la p	ró G	lv a	י פא				AA G	AC A	AG AT	LY CY	AA GC	ST GCC	1140	
										-, .		AL W	sn P	he G	ln A	sp L	is II	e G	ln Gl	GT GCC	380	
11	41 7	TT A	AC C	ርጥ ካር	CT C.															_		
3	81 F	he he	. D. N.		31 G	rr T	AC C	rt t	AT G	GT A	GT G	CT A	AC GO	GA A	אר אר	<u>፡</u> ሞ አረ				т ата		
		A.	311 M	rg Se	≥r Va	נו ונ	yr Le	u T	yr G	ly s	er A	la A	sn G1	1 V A			-A AA	C CC	TAT	T ATA	1200	
														· ,	311 SE	er Tr	r As	n Pr	0 Il	e Ile	400	
120	ol I	TA AG	A GO	SC GA	A AC	ב פר	ית די	·			•									I GGC		
40	)ı L	eu Ar	g G1	v Gl	11 Se	- 11		-A AA	rc cc	T C	CT GT	T TI	C TC	A GG	A TI	'A GA	T TA	T AL	~ aa-	T CCC		
			_			- AI	a In	I As	n Pr	0 P	ro Va	l Ph	e Se	r Gl	y Le	u As	D T12	- >		T GGC	1260	
336															•		y .	- AS	n Asi	u GIA	420	
126	1 T	AC CT	A TT	A AG	T AT	T GA	A GG	T GA	מד ד	T TC		<b>~</b> :								OSO 7		
.42	1 T)	T Le	u Le	u Se	r Il	e Gl	u GI	v be			. AA	I AT	T AA	A GA	T AT.	A GA	3 TT7	LAA.	ACT	GGG	1320	
		r Le					- <b></b> .	, vs	ріу	r Tr	P As	n Il	e Ly:	S As	p Il	e Gli	. Phe	Lvs	Thr	- G) v		
132	1 TC	יא איי													•			-4-		. Cly	440	
44:	1 0-	T AAJ r Lys	i GG	r ar	r GT7	CT:	r GA	AA:	T TC	T AA	T GG	ר אכי	י ת ח									
**.	- Se	r Lys	Gl3	/ Ile	≀ Va]	Let	a Asp	aA c	l Se	- 10	n Gl.		-	. 113	L AAA	AAC	CII	GTT	GTT	CAT	1380	
		r Lys					_				G1	, sei	Lys	Let	Lys	Asn	Leu	Val	Val	His	460	
1381	GA	I ATT	GCS	C2.																		
461	. As	T ATT	C1	GAA	GAA	. GCT	` ATT	CAC	TTC	G CG	GA1	GGA	TCT	AGC	י אא	, yyw						
		, 116	GIY	Glu	Glu	Ala	Ile	His	Leu	Arg	Ast	Glv	Sar	50-		AAI	AGT	ATA	GAT	GGT	1440	
		) Ile								•		,	261	ser	Asn	Asn	Ser	Ile	Asp	Gly	480	
1441	TG	ACT Thr	ATA	TAC	AAT	ארא		•														
481	Cys	Thr	Ile	Tier	200		GGI	AGA	ACT	AAA	CCI	GGT	TTT	GGT	GAA	GGT	TTL	ፖስጥ	CTN			
		Thr		- 7 -	ASI	inr	Gly	Arg	Thr	Lys	Pro	Gly	Phe	Glv	Glu	G1	1	-	GIA	GGC	1500	
1501												-	_	3		GIY	ren	IYI	Val	Gly	500	
1501	TCA	GAT Asp	AAA	GGA	CAA	CAT	GAC	سب لا	ጥስጥ	<b>~</b> · ·												
501	Ser	Asp	Lys	Gly	Gln	Hie	7	mt.	IAT	GAA	AGA	GCT	TGT	AAC	AAT	AAC	ACT	ATT	GAA	אאר	1560	
				•		3	vəb	TUL	Tyr	Glu	Arg	Ala	Cys	Asn	Asn	Asn	Th-	T1 ~				
1561	T		_															-16	OIU /	ASN	520	
	101	ACC Thr	GTT	GGA	CCC	TAA	GTA	ACA	GCA	CA A	ccc	<b>~</b>										
521	Cys	Thr	Val	Gly	Pro	Asn	Va)	Th-	۸۱-	C)		GIA	GAT	GTT	AAG	GAA	GGT ;	ACA	ATG /	AAC	1620	
		Thr				-			ur g	GIU	Gly	Val	Asp	Val	Lys	Glu .	Glv 1	Thr :	Met ?	965		
																	, ,	1		11Cr	540	

1621	1 ACT ATT ATA AGA AAT TGC GTG TTT TCT CCL CLL	
541	1 ACT ATT ATA AGA AAT TGC GTG TTT TCT GCA GAA GGA ATT TCA GGA GAA AAT AGC TCA GAT 1 Thr Ile Ile Arg Asn Cys Val Phe Ser Ala Glu Gly Ile Ser Gly Glu Asn Ser Ser Asp	1680
	of the Ser Ala Glu Gly Ile Ser Gly Glu Asn Ser Ser No.	
		560
1681	GCT TTT ATT GAT TTA AAA GGA GCC TAT GGT TTT GTA TAT	•
561	GCT TIT AIT GAT TTA AAA GGA GCC TAT GGT TTT GTA TAC AGA AAC ACG TTT AAT GTT GAT Ala Phe Ile Asp Leu Lys Gly Ala Tyr Gly Phe Val Tyr Arg Asn Thr Phe Asn Val Asp	1740
÷	and Tyr Gry Phe Val Tyr Arg Ash Thr Phe Ash Val Ash	580
		360
	GGT TCT GAA GTA ATA AAT ACT GGA GTA GAC TTT TTA GAT AGA GGT ACA GGA TTT AAT ACA	
201	Gly Ser Glu Val Ile Asn Thr Gly Val Asp Phe Leu Asp Arg Gly Thr Gly Phe Asn Thr	1800
	and ded Asp Arg Gly Thr Gly Phe Asn Thr	600 -
1801	GGT TTT AGA AAT GCA ATA TOTAL	
601	GGT TTT AGA AAT GCA ATA TTT GAA AAT ACA TAT AAC CTT GGC AGT AGA GCT TCA GAA ATT :	
	Gly Phe Arg Asn Ala Ile Phe Glu Asn Thr Tyr Asn Leu Gly Ser Arg Ala Ser Glu Ile	1860
		620
1861	THE GCT CGT AAA AAA CAA CCT TOT OF THE	
621	Ser Thr Ala Arg Lys Lys Gln Gly Ser Des Cl. CAC GTT TGG GAT AAT ATT AGA 1	1920
	The Gir The His Val Tro her her the	640
641	AAC CCT AAT TCT GTT GAT TTT CCA ATA AGT GAT GGT ACA GAA AAT CTA GTA AAT AAA TTC 1	
011	Asn Pro Asn Ser Val Asp Phe Pro Ile Ser Asp Gly Thr Glu Asn Leu Val Asn Lys Phe	980
	The Did Ash Let Val Ash Lys Phe	660
1981	TGC CCA GAT TGG AAT ATA GAA CCA TOT 110	
661	TGC CCA GAT TGG AAT ATA GAA CCA TGT AAT CCT GTA GAC GAA ACC AAC CAA GCA CCT ACA  Cys Pro Asp Trp Asn Ile Glu Pro Cys Asp Pro Yea	040
	Val Asp Glu Thr hen Cla has a	680
2041	ATA AGC TTC CTA TCT CCT GTT AAC AAT ATT ACT TTA GTT GAA GGT TAT AAT TTA CAA GTT 21  Lee Ser Phe Leu Ser Pro Val Asn Asn Lle The Leu Ser Pro Val Asn Lle The Leu Ser Pro Va	
981	The Ser Phe Leu Ser Pro Val Asn Asn He Thr Leu Val Charge 21	100
	Ile Ser Phe Leu Ser Pro Val Asn Asn Ile Thr Leu Val Glu Gly Tyr Asn Leu Gln Val	700
701 (	GAA GTT AAT GCT ACT GAT GCA GAT GGA ACT ATT GAT AAT GTA AAA CTT TAT ATA GAT AAC 21	60
	ASD ASD ASD Take Tour man	
		20
2161 )	AAT TTA GTT AGG CAA ATA AAT TCT ACT TCA TAT AAA TGG GGC CAT TCT GAT TCT CCA AAT 22:	
721 A	Asn Leu Val Arg Gln Ile Asn Ser Thr Ser Tur Luc T	20
	Asn Leu Val Arg Gln Ile Asn Ser Thr Ser Tyr Lys Trp Gly His Ser Asp Ser Pro Asn 74	40
741 T	ACA GAT GAA CTT AAT GGT CTT ACA GAA GGA ACT TAT ACC TTA AAA GCA ATT GCA ACT GAT 228	
	ord Gry The Tyr The Tage No. 73	
		50
2281 A	VAC GAC GGG GCT TCT ACA GAA ACG CAA TTT ACG TO ACG	
761 As	NAC GAC GGG GCT TCT ACA GAA ACG CAA TTT ACG TTA ACT GTA ATA ACA GAA CAA AGT CCG 234	0
	ASD ASD Gly Ala Ser Thr Glu Thr Gln Phe Thr Leu Thr Val Ile Thr Glu Gln Ser Pro 78	0
781 Se	CT GAG AAT TGT GAC TTT AAT ACA CCT TCT TCA ACT GGT TTA GAA GAT TTT GAC ATT AAA 240	
se	er Glu Asn Cys Asp Phe Asn Thr Pro Ser Ser Thr Gly Leu Glu Asp Phe Asp Ile Lys 800	0 .
		o .
2401 AA	AG TTT TCT AAC GTT TTT GAG TTA GGA TTA	
	AG TTT TCT AAC GTT TTT GAG TTA GGA TCT GGC GGA CCA TCT TTA AGT AAT TTA AAA ACA 2460	D
	•	-
	Figure 17c(continued)	

501	. Ly	s P	he S	er A	lsn '	Val	Phe	Gl u	Leu	Gly	' Se	r G1	y G	ly Pı	ro s	er	Let	1 Se	r As	n Le	eu L	ys Th	r 82
2461 821	Ph	T A(	T A	TT A le A	AT 1	rp ;	AAT Asn	TCG Ser	CAA Gln	TAC Tyr	AA: Asi	T GG n Gl	G TI Y Le	TA TA	T G	AA ln	TTI Phe	TC: Se:	A AT	A AA e As	C Ac	CA AA	C 252
2521 841	AAC	GG GG	T G	TA C	CT G	AT 7	YE I	TAT :	ATA Ile	AAT Asn	TTA	A AAI	A CC	A AA o Ly:	A A: s II	IT.	ACC Thr	TTT	CAC	3 TT	T AA e Ly	A AAT s Asn	2580
2581 861														PEC	AS	n F	he	Asp	Gly	Asp	Ty	TGG	800
2641 881													502	Dys	Th:	r A	sn .	Asn	Phe	Thr	Ile	Tyr	2700 900
											•			THE	PIC	) Se	er )	lsn	Gln	Ile	Ser	AAA Lys	2760 920
											•		- 7 -	CCT Pro	Asn	Pr	о д	la 1	Leu .	Asp	GAA Glu	ACT Thr	2820 940
821 )	ATT :	LTT	GTG	AGC	GCI	GA	(C)	ר כא						GTG   Val					FT 21	970 956			

Figure 17d(continued)

## Figure No. 180 Pyrococcus furiosus VC1(7EG1)

lea	der	· se	đu	ence	e: a	mino	aci	ds 1	-24											
					• .	•	1				7		3	6			_		_	
5 '	AT	G A	(GC	AAC	AA E	A AA	G TT	C GT	C AT	C GI	A TO	T AT	יייים דייים כו	א אכ	LA AI	4				4
	Me	t S	er	Lys	Ly	s Ly	s Ph	e Va	1 11	e Va	l Se	r II	le Le	11 Th	r Il	e te	1 11	A GI	A CA	G
														· 4 11.		e ne	r re	u Va	1 G1	n
				63			7:	-	•		1		9	0		9	<u> </u>			
•	GC	A A	TA	TAT	TT	r GT	A GA	A AA	G TA	T CA	T AC	C TO	T GA	G GZ	AA D.	G TO		m ===	10	
	Al	a I	le	Tyr	Phe	≥ Va	l Gli	Ly:	s Ty	r Hi	s Th	r Se	r Gl	u As	p Ly	5 IC	- ML	r re	A AA	T
															p by	3 JE.	L III	r Se	r As:	n
				117			126	5		13:	5		14	a						
	ACC	Т	CA.	TCT	ACA	CC	ccc	CA	AC	A AC	A CT	T TC	C AC	ר אר	C AA	151		~	16:	_
	Thi	S	eŗ	Ser	Thr	Pro	Pro	Glr	Thi	Th	Le	u Se	r Th:	r Th	r Ly	5 Wal	t Cit	- AA	G AT	r
															L Dy.	s val	. TE	ı Ly	2 116	•
				171			180	)		189	,		196	a		202				_
	AGA	T	AC	CCT	GAT	GAC	GGI	GAG	TGG	CC	. GGZ	A GC	r cci	ייד ב	r gat	207			216	
	Arg	T	YI	Pro	Asp	Asp	Gly	Glu	Trp	Pro	Gly	Ala	a Pro	o Ile	Asp	LAMO	DA.		r GAI	
											-					, mys	ASI	61)	Asp	)
				225			234			243			252	,		263				
	GGG	LA.	/C	CCA	GAA	TTC	TAC	ATT	GAA	ATA	AAC	CT	TGG	אא	TTA:	261	<b>5</b> 2 0		270	
	Gly	Αε	n	Pro	Glu	Phe	Tyr	Ile	Glu	Ile	Asn	Lev	1 Trp	Asn	Ile	Ten	AAI	חום	ACT	
																Dea	ASU	, ATG	Inr	
			:	279			288			297			306			315			224	
	GGA	TI	T (	GCT	GAG	ATG	ACG	TAC	AAT	TTA	ACC	AGC	GGC	GTC	CTT	C) C	The	CTC	324	
	Gly	Ph	e i	Ala	Glu	Met	Thr	Tyr	Asn	Leu	Thr	Ser	Gly	Val	Leu	Hie	Tare	37-1	CAA	
													. •				- 7 -	VAI	. 6111	
				333	•		342			351			360			369	•			
•	CAA	CT	T	BAC	AAC	ATT	GTC	TTG	AGG	GAT	AGA	AGT	AAT	TGG	GTG	ראידי	GGN	TD C	378	
•	Gln	Le	u I	4sp	Asn	Ile	Val	Leu	Arg	Asp	Arg	Ser	Asn	Trp	Val	His	GIV	THE	Pro	
														•			CLy	171	PLO	
				87			396			405			414.			423			420	
(	SAA	AT	A I	TC	TAT	GGA	AAC	AAG	CCA	TGG	AAT	GCA	A A C	TAC	GCA		מאת	ccc	432	
(	Slu	Il	e P	he '	Tyr	Gly	Asn	Lys	Pro	Trp	Asn	Ala	Asn	Tvr	Ala	Thr	DED	Clar	Dec	
														- 4			nap	GIY	ET.O.	
				41			450			459			468							
A	TA	CCI	T	TA (	ccc .	AGT	AAA	GTT	TCA	AAC	CTA	ACA	CAC	TTC	TAT	477	200		486	
Į	le	Pro	L	eu 1	Pro	Ser	Lys	Val	Ser	Asn	Leu	Thr	Asp	Phe	Tyr	Len	ACA	ATC	TCC	
													F		- 7 1	⊒eu	Tur	TIE	ser	

TAT AAA CTT GAG CCC AAG AAC GGC CTG CCA ATT AAC TTC GCA ATA GAA TCC TGG
Tyr Lys Leu Glu Pro Lys Asn Gly Leu Pro Ile Asn Phe Ala Ile Glu Ser Trp

549 558 567 576 585 594

TTA ACG AGA GAA GCT TGG AGA ACA ACA GGA ATT AAC AGC GAT GAG CAA GAA GTA

Leu Thr Arg Glu Ala Trp Arg Thr Thr Gly Ile Asn Ser Asp Glu Gln Glu Val

603 612 621 630 639 648

ATG ATA TGG ATT TAC TAT GAC GGA TTA CAA CCG GCT GGC TCC AAA GTT AAG GAG

Met Ile Trp Ile Tyr Tyr Asp Gly Leu Gln Pro Ala Gly Ser Lys Val Lys Glu

657 666 675 684 693 702
ATT GTA GTC CCA ATA ATA GTT AAC GGA ACA CCA GTA AAT GCT ACA TTT GAA GTA
Ile Val Val Pro Ile Ile Val Asn Gly Thr Pro Val Asn Ala Thr Phe Glu Val

TGG AAG GCA AAC ATT GGT TGG GAG TAT GTT GCA TTT AGA ATA AAG ACC CCA ATC
TTP Lys Ala Asn Ile Gly Trp Glu Tyr Val Ala Phe Arg Ile Lys Thr Pro Ile

765 774 783 792 801 810
AAA GAG GGA ACA GTG ACA ATT CCA TAC GGA GCA TTT ATA AGT GTT GCA GCC AAC
Lys Glu Gly Thr Val Thr Ile Pro Tyr Gly Ala Phe Ile Ser Val Ala Ala Asn

819 828 837 846 855 864
ATT TCA AGC TTA CCA AAT TAC ACA GAA CTT TAC TTA GAG GAC GTG GAG ATT GGA
Ile Ser Ser Leu Pro Asn Tyr Thr Glu Leu Tyr Leu Glu Asp Val Glu Ile Gly

ACT GAG TTT GGA ACG CCA AGC ACT ACC TCC GCC CAC CTA GAG TGG TGG ATC ACA
Thr Glu Phe Gly Thr Pro Ser Thr Thr Ser Ala His Leu Glu Trp Trp Ile Thr

927 936 945 954

AAC ATA ACA CTA ACT CCT CTA GAT AGA CCT CTT ATT TCC TAA 3'

Asn Ile Thr Leu Thr Pro Leu Asp Arg Pro Leu Ile Ser *

Figure 18b(continued)

International application No. PCT/US97/22623

	SSIFICATION OF SUBJECT MATTER											
	:C07H 21/04; C12N 1/20, 1/14, 5/00, 9/38, 9/42; (2435/207, 209, 252.3, 254.11, 274, 275, 320.1, 32:											
	to International Patent Classification (IPC) or to both											
B. FIEL	LDS SEARCHED											
Minimum d	locumentation searched (classification system followe	d by classification symbols)										
U.S. :	435/207, 209, 252.3, 254.11, 274, 275, 320.1, 325	; 536/23.2										
Documenta	tion searched other than minimum documentation to the	extent that such documents are included	in the fields searched									
ļ												
Electronic of	data base consulted during the international search (na	ame of data base and, where practicable	e, search terms used)									
Please Se	e Extra Sheet.	•										
	·											
C. DOC	UMENTS CONSIDERED TO BE RELEVANT											
Category*	Citation of document, with indication, where ap	ppropriate, of the relevant passages	Relevant to claim No.									
x	GRABNITZ et al. Structure of the f	B-Glucosidase Gene bglA of	1-3, 5									
	Clostridium thermocellum: Sequence Analysis Reveals a Superfamily species II											
A												
	Hydrolase. Eur. J. Biochem. Septemb	ber 1991, Vol. 200, No. 2,	4, 6-11									
· .	pages 301-309, see entire document.											
$ \mathbf{x} $	VOORHORST et al. Characterization	of the celB Gene Coding for	1-3, 5									
	β-Glucosidase from the Hyperthermo	<u> </u>	species I and III									
A	furiosus and Its Expression and Site-Dir											
	coli. J. Bacteriol. December 1995, Vo		4, 6-11									
	7111, see entire document.	· -										
	·	:										
			•									
Furti	ner documents are listed in the continuation of Box C	C. See patent family annex.										
· ·	ecial categories of cited documents:	"T" later document published after the inte date and not in conflict with the appl										
	cument defining the general state of the art which is not considered be of particular relevance	the principle or theory underlying the										
i	rlier document published on or after the international filing date	"X" document of particular relevance; the considered novel or cannot be considered.										
cit	cument which may throw doubts on priority claim(s) or which is ed to establish the publication date of another citation or other	when the document is taken alone										
	ecist reason (as specified)  cument referring to an oral disclosure, use, exhibition or other	"Y" document of particular relevance; the considered to involve an inventive combined with one or more other suc-	step when the document is									
me	nems .	being obvious to a person skilled in	the art									
the	ecument published prior to the international filing date but later than e priority date claimed	*& document member of the same paten										
Date of the	actual completion of the international search	Date of mailing of the international se	aren report									
26 MARG	CH 1998	<b>2</b> 1 APR 1998										
Name and I	mailing address of the ISA/US	Authorized officer	h									
Box PCT	oner of Patents and Trademarks	LISA J. HOBBS, PH.D.	NV4/									
Washington	n, D.C. 20231	Telephone No. (703) 308-0106	" Gir									

International application No. PCT/US97/22623

Box I	Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)
This into	ernational report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:
I	Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:
2.	Claims Nos.: because they relate to parts of the international application that do not comply with the prescribed requirements to such
	an extent that no meaningful international search can be carried out, specifically:
3. 🗀	Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).
Box II	Observations where unity of invention is lacking (Continuation of item 2 of first sheet)
	mational Searching Authority found multiple inventions in this international application, as follows:
Pl	case See Extra Sheet.
· 🗀	As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
	As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
). X	As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:  11, species I-III
<del>-</del> ,	AT, Species I-III
· 🗀	No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
lemark (	on Protest
	No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet(1))(July 1992)*

International application No. PCT/US97/22623

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)
This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:
1. Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:
2. Claims Nos.: because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).
Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)
This International Searching Authority found multiple inventions in this international application, as follows:
Please See Extra Sheet.
1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. X As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
1-11, species I-III
4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
Remark on Protest  The additional search fees were accompanied by the applicant's protest.
No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet(1))(July 1992)*

International application No. PCT/US97/22623

#### **B. FIELDS SEARCHED**

Electronic data bases consulted (Name of data base and where practicable terms used):

APS and STN (Bioscience and Patent Indexes): Desulfurococc##, Staphylotherm##, Thermatoga, galactosidase#, glucosidase#, beta galactosidase#, beta glucosidase#. Genbank, EMBL, ESTs1-4, STS, N-Geneseq: Seq. ID Nos.: 1-3 and A-Geneseq, PIR, Swissprot: Seq ID Nos.: 15-17.

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING This ISA found multiple inventions as follows:

This application contains claims directed to more than one species of the generic invention. These species are deemed to lack Unity of Invention because they are not so linked as to form a single inventive concept under PCT Rule 13.1. The species are as follows: there are 18 distinct enzymes disclosed in the description, as enumerated in Figs. 1-18 and Table 1.

The claims are deemed to correspond to the species listed above in the following manner: while all the claims form one Group for examination, each of the claims is generic to the 18 enzyme species disclosed.

The species listed above do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, the species lack the same or corresponding special technical features for the following reasons: each enzyme is a different product, thus has the special technical feature of the recited enzyme, which the other species lack.

Form PCT/ISA/210 (extra sheet)(July 1992) *

International application No. PCT/US97/22623

#### **B. FIELDS SEARCHED**

Electronic data bases consulted (Name of data base and where practicable terms used):

APS and STN (Bioscience and Patent Indexes): Desulfurococc##, Staphylotherm##, Thermatoga, galactosidase#, glucosidase#, beta galactosidase#, beta glucosidase#. Genbank, EMBL, ESTs1-4, STS, N-Geneseq: Seq. ID Nos.: 1-3 and A-Geneseq, PIR, Swissprot: Seq ID Nos.: 15-17.

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING This ISA found multiple inventions as follows:

This application contains claims directed to more than one species of the generic invention. These species are deemed to lack Unity of Invention because they are not so linked as to form a single inventive concept under PCT Rule 13.1. The species are as follows: there are 18 distinct enzymes disclosed in the description, as enumerated in Figs. 1-18 and Table 1.

The claims are deemed to correspond to the species listed above in the following manner: while all the claims form one Group for examination, each of the claims is generic to the 18 enzyme species disclosed.

The species listed above do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, the species lack the same or corresponding special technical features for the following reasons: each enzyme is a different product, thus has the special technical feature of the recited enzyme, which the other species lack.

Form PCT/ISA/210 (extra sheet)(July 1992) *

·			•
			•
			- ·
	· · · · · · · · · · · · · · · · · · ·		
		•	
			· ·
		•	
			·
			•
			٠.